

Copyright GenCore version 5.1.6 (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw mode1

Run on: March 9, 2005, 02:23:54 ; Search time 1671 Seconds (without alignments)

666.948 Million cell updates/sec

Title: US-09-835-976B-115

Perfect score: 23

Sequence: 1 gtgtcttaaggagacaccacagc 23

Scoring table: IDENTITY NUC Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0% Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
 1: gb:ba:/*
 2: gb:htg:/*
 3: gb:ln:/*
 4: gb:om:/*
 5: gb:ov:/*
 6: gb:pat:/*
 7: gb:ph:/*
 8: gb:pl:/*
 9: gb:pr:/*
 10: gb:ro:/*
 11: gb:sts:/*
 12: gb:sy:/*
 13: gb:un:/*
 14: gb:vi:/*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
C 1	23	100.0	63808	2	AC090554		AC090554 Homo sapi
C 2	23	100.0	150791	9	AL162458		AL162458 Human DNA
C 3	23	100.0	185931	2	AL13554		AL13554 Homo sapi
C 4	19.8	86.1	63808	2	AC090554		AC090554 Homo sapi
C 5	19.8	86.1	178050	2	AC142988		AC142988 Macaca mulatta DNA
C 6	19	82.6	174956	10	AC131802		AC131802 Mus musculus
C 7	19	82.6	197519	2	AC15553		AC15553 Mus musculus
C 8	19	82.6	299377	2	AC127238		AC127238 Mus musculus
C 9	18.8	81.7	100711	9	AY62109		AY62109 Homo sapi
C 10	18.8	81.7	128625	10	AL837520		AL837520 Mouse DNA
C 11	18.8	81.7	149288	9	AC113144		AC113144 Homo sapi
C 12	18.8	81.7	162246	10	BX00428		BX00428 Mouse DNA
C 13	18.8	81.7	185212	2	BX572088		BX572088 Mus musculus
C 14	18.8	81.7	193572	10	AC140268		AC140268 Mus musculus
C 15	18.8	81.7	194490	10	AL845309		AL845309 Mouse DNA
C 16	18.8	81.7	200441	2	AC113895		AC113895 Rattus norvegicus
C 17	18.8	81.7	202992	9	AC068418		AC068418 Homo sapi
C 18	18.8	81.7	248196	2	AC117160		AC117160 Rattus norvegicus
C 19	18.8	81.7	277898	2	AC117083		AC117083 Rattus norvegicus

ALIGNMENTS

RESULT	AC090554	C	LOCUS	AC090554	C	AC090554	Hom sapiens	Chromosome 17	63808 bp	DNA	linear	HTG 03-MAR-2001
			DEFINITION	DEFINITION			SEQUENCE SAMPLING					Continuation (2 of 2)
			ACCESSION	AC090554								AC084837 Homo sapi
			VERSION	AC090554.1								AC022188 Homo sapi
			KEYWORDS									AC139542 Oryza sat
			SOURCE									AL138954 Human DNA
			ORGANISM									AC132255 Mus musculus
			REFERENCE									AC112923 Mus musculus
			AUTHORS									Birren, B., Linton, L., Nusbaum, C. and Lander, E.
			JOURNAL									Unpublished
			REFERENCE									2 (bases 1 to 63808)
			AUTHORS									Birren, B., Linton, L., Nusbaum, C. and Lander, E.
			JOURNAL									Unpublished
			REFERENCE									1 (bases 1 to 63808)
			AUTHORS									Birren, B., Linton, L., Nusbaum, C. and Lander, E.
			JOURNAL									Unpublished
			REFERENCE									2 (bases 1 to 63808)
			AUTHORS									Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, I., Boukhalter, B., Brown, A., Camarata, O., Campopiano, A., Choepel, Y., Colangelo, M., Grand-Pierre, N., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hafner, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T., Lehoczky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McKernan, K., McEwan, P., McPhee, R., McElrath, J., Meneus, I., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollard, V., Raymond, C., Rieback, M., Riley, R., Riese, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Souza, C., Spencer, B., Stange, J., Stojanovic, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
			JOURNAL									Direct Submission
			COMMENT									Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute / MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: Li0899

Center clone name: 191_D_16

 * NOTE: This record contains 75 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

1 724: contig of 724 bp in length

* 725 824: gap of 100 bp

* 825 1636: contig of 812 bp in length

* 1637 1736: gap of 100 bp

* 1737 2433: contig of 697 bp in length

* 2434 2533: gap of 100 bp

* 2534 3256: contig of 723 bp in length

* 3257 3356: gap of 100 bp

* 3357 4122: contig of 766 bp in length

* 4123 4222: gap of 100 bp

* 4223 4966: contig of 744 bp in length

* 4967 5066: gap of 100 bp

* 5067 5818: contig of 752 bp in length

* 5819 5918: gap of 100 bp

* 5919 6623: Contig of 705 bp in length

* 6624 6723: gap of 100 bp

* 6724 7535: contig of 812 bp in length

* 7536 7635: gap of 100 bp

* 7636 8413: contig of 778 bp in length

* 8414 8513: gap of 100 bp

* 8514 9224: contig of 711 bp in length

* 9225 9324: gap of 100 bp

* 9325 10046: contig of 722 bp in length

* 10047 10146: gap of 100 bp

* 10147 10893: contig of 747 bp in length

* 10894 10993: gap of 100 bp

* 10994 11760: contig of 767 bp in length

* 11761 11860: gap of 100 bp

* 11861 12595: contig of 735 bp in length

* 12596 12695: gap of 100 bp

* 12696 13435: contig of 740 bp in length

* 13436 13535: gap of 100 bp

* 13536 14320: contig of 785 bp in length

* 14321 14420: gap of 100 bp

* 14421 15142: contig of 722 bp in length

* 15143 15242: gap of 100 bp

* 15243 16023: contig of 781 bp in length

* 16024 16123: gap of 100 bp

* 16124 16881: contig of 758 bp in length

* 16882 16981: gap of 100 bp

* 16982 17728: contig of 747 bp in length

* 17729 17828: gap of 100 bp

* 17829 18525: contig of 697 bp in length

* 18526 18625: gap of 100 bp

* 18626 19381: contig of 756 bp in length

* 19382 19481: gap of 100 bp

* 19482 20247: contig of 766 bp in length

* 20248 20347: gap of 100 bp

* 20348 21099: contig of 752 bp in length

* 21100 21199: gap of 100 bp

* 21190 21200: contig of 731 bp in length

 21331 22030: gap of 100 bp
 22331 22739: contig of 769 bp in length
 22899: gap of 100 bp
 22900 23669: contig of 750 bp in length
 23650 23749: gap of 100 bp
 23740 24473: contig of 724 bp in length
 24474 24573: gap of 100 bp
 24574 25337: contig of 784 bp in length
 25338 25457: gap of 100 bp
 25458 26221: contig of 764 bp in length
 26222 26311: gap of 100 bp
 26312 27091: contig of 770 bp in length
 27192 27927: contig of 736 bp in length
 27928 28027: gap of 100 bp
 28028 28732: contig of 755 bp in length
 28733 28883: gap of 100 bp
 28883 29601: contig of 719 bp in length
 29602 29701: gap of 100 bp
 29702 30481: contig of 780 bp in length
 30482 30581: gap of 100 bp
 30582 31361: contig of 780 bp in length
 31362 31451: gap of 100 bp
 31452 32117: contig of 686 bp in length
 32148 32247: gap of 100 bp
 32248 33045: contig of 778 bp in length
 33026 33126: gap of 100 bp
 33126 33839: contig of 714 bp in length
 33940 33949: gap of 100 bp
 33940 34680: contig of 741 bp in length
 34681 34780: gap of 100 bp
 34781 35528: contig of 748 bp in length
 35529 35638: gap of 100 bp
 35639 36401: contig of 773 bp in length
 36402 36501: gap of 100 bp
 36502 37266: contig of 765 bp in length
 37267 37356: gap of 100 bp
 37356 37367: contig of 754 bp in length
 38120 38220: gap of 100 bp
 38221 38952: contig of 732 bp in length
 38953 39051: gap of 100 bp
 39053 39812: contig of 760 bp in length
 39813 39912: gap of 100 bp
 39913 40642: contig of 737 bp in length
 40650 40743: gap of 100 bp
 40750 41524: contig of 775 bp in length
 41525 41624: gap of 100 bp
 41625 42384: contig of 760 bp in length
 42385 42484: gap of 100 bp
 42485 43241: contig of 757 bp in length
 43242 43341: gap of 100 bp
 43342 44105: contig of 764 bp in length
 44106 44205: gap of 100 bp
 44206 44977: contig of 772 bp in length
 44978 45079: gap of 100 bp
 45079 45855: contig of 778 bp in length
 45856 45955: gap of 100 bp
 45956 46731: contig of 751 bp in length
 46732 46831: gap of 100 bp
 46832 47504: contig of 673 bp in length
 47505 47604: gap of 100 bp
 47605 48379: contig of 759 bp in length
 48380 48479: gap of 100 bp
 48480 49230: contig of 743 bp in length
 49231 49330: gap of 100 bp
 49331 50130: contig of 800 bp in length
 50131 50230: gap of 100 bp
 50231 50989: contig of 759 bp in length
 50990 51090: gap of 100 bp
 51090 51832: contig of 743 bp in length
 51833 51932: gap of 100 bp
 51933 52684: contig of 752 bp in length
 52685 52784: gap of 100 bp

/mol type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="RPCI-91G13"
 /clone_id="RPCI-1"
 1..34065
 /note="assembly_fragment:02324
 fragment_chain:1
 clone_end:17
 misc_feature
 34166..34941
 /note="assembly_fragment:01244
 fragment_chain:1
 clone_end:SP6
 vector_side:right"
 7942..141187
 /note="assembly_fragment:01611"
 141288..185931
 /note="assembly_fragment:02960"
 misc_feature
 107433..107455

Query Match Score 23; DB 2; Length 185931;
 Best Local Similarity 100.0%; Pred. No. 0.42; 0; Mismatches 0; Indels 0; Gaps 0;
 Matches 23; Conservative 0; Gapopen 0;

ORIGIN
 AC090554 AC090554 Homo sapiens chromosome 17 clone RP1-191D16 map 17, LINEAR HTG 03-MAR-2001
 DEFINITION SEQUENCE SAMPLING.

ACCESSION AC090554
 KEYWORDS GI:13194277
 HTGS; HTGS_PHAEB0
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 (bases 1 to 63808)

REFERENCE Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome 17, clone RP1-191D16
 JOURNAL Unpublished
 JOURNAL 2 (bases 1 to 63808)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Venter, J., O'Farrell, P., O'Neil, D., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamazaro, R., Landers, T., Lehoczky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Margis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeeters, R., Meldrim, J., Menes, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C., H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Pankhania, P., Pierre, N., Pollar, V., Raymond, C., Retta, R., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Sognez, C., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zimmer, L., Zimmer, A. and Zody, M.

JOURNAL Direct Submission
 JOURNAL Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 Genome Center:
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-beg.wi.mit.edu>
 Contact: sequence_submission@genome.wi.mit.edu
 Project Information
 Center project name: Li1899
 Center clone name: 191_D_16

NOTE: This record contains 75 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

724: contig of 724 bp in length
 824: gap of 100 bp
 1636: contig of 812 bp in length
 1637: gap of 100 bp
 1737: contig of 697 bp in length
 2433: gap of 100 bp
 2533: gap of 100 bp
 2534: contig of 723 bp in length
 3234: gap of 100 bp
 3257: gap of 100 bp
 3257: contig of 766 bp in length
 4122: gap of 100 bp
 4123: contig of 744 bp in length
 4223: gap of 100 bp
 4966: contig of 744 bp in length
 5067: gap of 100 bp
 5878: contig of 752 bp in length
 5918: gap of 100 bp
 5919: contig of 705 bp in length
 6633: gap of 100 bp
 6724: gap of 100 bp
 6724: contig of 812 bp in length
 7536: gap of 100 bp
 7636: contig of 778 bp in length
 8414: gap of 100 bp
 8514: contig of 711 bp in length
 9325: gap of 100 bp
 10016: contig of 722 bp in length
 10146: gap of 100 bp
 1047: contig of 747 bp in length
 10833: gap of 100 bp
 10994: contig of 767 bp in length
 11094: gap of 100 bp
 11880: gap of 100 bp
 11861: contig of 735 bp in length
 12596: gap of 100 bp
 13455: contig of 740 bp in length
 13436: 13436
 14320: contig of 785 bp in length
 14421: 14421
 15142: contig of 722 bp in length
 15242: gap of 100 bp
 15443: 15443
 16033: contig of 781 bp in length
 16123: gap of 100 bp
 16124: contig of 758 bp in length
 16881: gap of 100 bp
 16882: 16882
 17788: contig of 747 bp in length
 17888: gap of 100 bp
 17929: 17929
 18555: contig of 697 bp in length
 18826: gap of 100 bp
 19311: contig of 756 bp in length
 19411: gap of 100 bp
 19482: 19482
 20247: contig of 766 bp in length
 20347: gap of 100 bp
 20348: 20348
 21099: contig of 752 bp in length
 21100: gap of 100 bp
 21190: contig of 731 bp in length
 22030: gap of 100 bp

21931:

22031	22799:	contig	of	769	bp	in	length
22890	22899:	gap	of	100	bp		
22990	23649:	contig	of	750	bp	in	length
23650	23749:	gap	of	100	bp		
23750	24473:	contig	of	724	bp	in	length
24474	24573:	gap	of	100	bp		
24574	25357:	contig	of	784	bp	in	length
25358	25458:	gap	of	100	bp		
25458	26221:	contig	of	764	bp	in	length
26222	26321:	gap	of	100	bp		
26322	27091:	contig	of	770	bp	in	length
27092	27191:	gap	of	100	bp		
27192	27922:	contig	of	736	bp	in	length
27928	28022:	gap	of	100	bp		
28029	28782:	contig	of	755	bp	in	length
28783	28882:	gap	of	100	bp		
28883	29601:	contig	of	719	bp	in	length
29602	29701:	gap	of	100	bp		
29702	29792:	contig	of	736	bp	in	length
30481	30481:	gap	of	100	bp		
30505	31361:	contig	of	780	bp	in	length
31362	31461:	gap	of	100	bp		
31462	32147:	contig	of	686	bp	in	length
32148	32247:	gap	of	100	bp		
32248	33025:	contig	of	778	bp	in	length
33026	33125:	gap	of	100	bp		
33126	33829:	contig	of	714	bp	in	length
33830	33939:	gap	of	100	bp		
33940	34680:	contig	of	741	bp	in	length
34681	34780:	gap	of	100	bp		
34781	35528:	contig	of	748	bp	in	length
35529	35628:	gap	of	100	bp		
35629	36401:	contig	of	773	bp	in	length
36402	36501:	gap	of	100	bp		
36502	37266:	contig	of	765	bp	in	length
37267	37366:	gap	of	100	bp		
37367	38120:	contig	of	754	bp	in	length
38121	38220:	gap	of	100	bp		
38221	38952:	contig	of	732	bp	in	length
38953	39052:	gap	of	100	bp		
39053	39812:	contig	of	760	bp	in	length
39813	39912:	gap	of	100	bp		
39913	40649:	contig	of	737	bp	in	length
40650	40749:	gap	of	100	bp		
40750	41524:	contig	of	775	bp	in	length
41525	41624:	gap	of	100	bp		
41625	42384:	contig	of	760	bp	in	length
42385	42484:	gap	of	100	bp		
42485:	43244:	contig	of	757	bp	in	length
43245	43341:	gap	of	100	bp		
43342	44105:	contig	of	764	bp	in	length
44106	44205:	gap	of	100	bp		
44206	44977:	contig	of	772	bp	in	length
44978	45078:	gap	of	100	bp		
45078	45855:	contig	of	778	bp	in	length
45856	45955:	gap	of	100	bp		
45956	46731:	contig	of	776	bp	in	length
46732	46831:	gap	of	100	bp		
46832	47504:	contig	of	673	bp	in	length
47505	47604:	gap	of	100	bp		
47605:	48375:	contig	of	775	bp	in	length
48376	48479:	gap	of	100	bp		
48480	49221:	contig	of	751	bp	in	length
49221	49330:	gap	of	100	bp		
49331	50130:	contig	of	800	bp	in	length
50131	50231:	gap	of	100	bp		
50232:	50989:	contig	of	759	bp	in	length
50990	51089:	gap	of	100	bp		
51090	51832:	contig	of	743	bp	in	length
51833	51932:	gap	of	100	bp		
51933	52684:	contig	of	752	bp	in	length
52685	53585:	gap	of	100	bp		
53586	54275:	contig	of	801	bp	in	length
54275	55275:	gap	of	100	bp		

Nguyen, A., Nguyen, N., Nickerson, B., Nwokonkwo, S., Oguh, M., Okuongu, G., Oreaganye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.-L., Quiles, M., Ren, Y., Rivas, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shooobtari, N., Sisson, I., Sodergren, E., Soniak, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Taylor, T., Telford, B., Vinson, R., Wang, Q., Wang, S., Ward, Moore, S., Warren, R., Warren, R., Washington, C., Watling, S., Williamson, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y.-F., Zhou, J., Zorrilla, S., Kucherlapati, R., Weinstock, G. and Gibbs, R.	Db	169950	GTGTCCTAGGAGACCAAGC 169872
RESULT 6			
AC131802	AC131802	174956	DNA linear RP24-542I6, complete sequence.
LOCUS	Mus musculus chromosome 7 cClone RP24-542I6		
DEFINITION			
ACCESSION	AC131802		
VERSION	GI:53380911		
KEYWORDS	HTG		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Muridae; Murinae; Mus.			
REFERENCE 1	(bases 1 to 174956)		
AUTHORS Wilson, R. K.			
TITLE The sequence of <i>Mus musculus</i> clone			
JOURNAL Unpublished			
REFERENCE 2	(bases 1 to 174956)		
AUTHORS McPherson, J.D. and Waterston, R.H.			
TITLE Direct Submission			
JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA			
REFERENCE 3	(bases 1 to 174956)		
AUTHORS Wilson, R. K.			
TITLE Direct Submission			
JOURNAL Submitted (30-APR-2004) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA			
REFERENCE 4	(bases 1 to 174956)		
AUTHORS Wilson, R. K.			
TITLE Direct Submission			
JOURNAL Submitted (01-OCT-2004) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA			
COMMENT On Oct 1, 2004 this sequence version replaced gi:46879428.			
FEATURES			
source			
1. 174956			
Organism: "Mus musculus"			
/mol_type="genomic DNA"			
/db_xref="taxon:10909"			
/chromosome="7"			
/clone="RP24-542I6"			
ORIGIN			
Query Match	82.6%	Score 19;	DB 10;
Best Local Similarity	100.0%	Pred. No. 53;	
Matches	19;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1	GTGTCCTAGGAGACCA 19	
Db	138192	GTGTCCTAGGAGACCA 138210	
RESULT 7			
AC145553/c	AC145553	197519	DNA linear HTG 30-APR-2004
LOCUS	Mus musculus chromosome 7 cClone RP23-106M1, WORKING DRAFT SEQUENCE,		
DEFINITION	6 unordered pieces.		
ACCESSION	AC145553		
VERSION	GI:46879441		
KEYWORDS	HTG; HTGS PHASB1; HTGS DRAFT; HTGS_ACTIVEFIN.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
misc_feature			
Query Match	86.1%	Score 19.8;	DB 2;
Best Local Similarity	91.3%	Pred. No. 20;	Length 178050;
Matches	21;	Conservative 0; Mismatches 2;	Indels 0; Gaps 0;
CONFIDENCE	_0.85"		
ORIGIN			
Qy	1	GTGTCCTAGGAGACCAAGC 23	

ORIGIN

EFFECTIVE 1. (bases 1 to 197519) Query Match 82.6%; Score 19; DB 2; Length 197519;

AUTHORS Wilson, R.K. Best Local Similarity 100.0%; Pred. No. 52;

JOURNAL The sequence of *Mus musculus* clone Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

EFFECTIVE 2. (bases 1 to 197519) Qy 1 GTCCTTAAGGAGACCA 19

AUTHORS Wilson, R.K. Db 21310 GTGCTTAAGGAGACCA 21292

JOURNAL Unpublished (19-JUL-2003) Genome Sequencing Center, 4444 Forest Park

JOURNAL Parkway, St. Louis, MO 63108, USA

JOURNAL On Apr 30, 2004 this sequence version replaced gi:40737111.

COMMENT

Center: Washington Genome Sequencing Center, 4444 Forest Park

Center code: WUGSC

Web site: <http://genome.wustl.edu>

Contact: submissions@watson.wustl.edu

Project Information

Center project name: M_PA0106M01

Summary Statistics

Sequencing vector: M13, 0%
 Sequencing chemistry: Dye-primer ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 195074 bases at least Q40
 Consensus quality: 195598 bases at least Q30
 Consensus quality: 195933 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1. 1505: contig of 1505 bp in length
 1506: gap of unknown length
 1606: 5809: contig of 4104 bp in length
 5810: 5909: gap of unknown length
 5910: 12821: contig of 6518 bp in length
 12828: 12937: gap of unknown length
 12928: 32221: contig of 19294 bp in length
 32221: 32321: gap of unknown length
 32222: 89970: contig of 57649 bp in length
 89971: 90070: gap of unknown length
 90071: 197519: contig of 107449 bp in length.

Location/Qualifiers

1. 1.197519
 /organism="Mus musculus"
 /mol type="Genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP23-106M1"

misc_feature

1. 1.1505
 /note="assembly_name:Contig17"
 1606. .5809
 /note="assembly_name:Contig26"
 5910. .12827
 5910. .12937: contig of 1137 bp in length
 1138. 1237: gap of unknown length
 1238. 2336: contig of 1099 bp in length
 2337. 2436: gap of unknown length
 2437. 3705: contig of 1269 bp in length
 3706. 3805: gap of unknown length
 3805. 11575: gap of 7770 bp in length
 11575. 11675: gap of unknown length
 11575. 11575: gap of 7770 bp in length
 11575. 11575: gap of unknown length

source

misc_feature

1. 1.1505
 /note="assembly_name:Contig27"
 12928. .32221
 /note="assembly_name:Contig28"
 32222. .89970
 /note="assembly_name:Contig29"
 9071. .197519
 /note="assembly_name:Contig30"

11676 17379: contig of 5704 bp in length
 * 17380 17479: gap of unknown length
 * 17480 26373: contig of 8893 bp in length
 * 17480 26472: gap of unknown length
 * 17480 26473: contig of 19647 bp in length
 * 17480 45919: gap of unknown length
 * 17480 46020: contig of 26280 bp in length
 * 17480 72300: gap of unknown length
 * 17480 72399: contig of 32426 bp in length
 * 17480 104825: gap of unknown length
 * 174826 104925: gap of unknown length
 * 174826 104926: contig of 69659 bp in length
 * 174858 174684: gap of unknown length
 * 174685 237121: contig of 62437 bp in length
 * 237122 237221: gap of unknown length
 * 237222 299377: contig of 62156 bp in length.

FEATURES source
 1. .299377
 Location_Qualifiers
 1. .299377
 /organism="Mus musculus"
 /mol_type="Genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="7"
 /clone="RP24-378P20"
 /clone="RP24-378P20"
 /note="assembly_name:Contig34"
 misc_feature
 /note="assembly_name:Contig37"
 misc_feature
 /note="assembly_name:Contig39"
 misc_feature
 /note="assembly_name:Contig45"
 misc_feature
 /note="assembly_name:Contig46"
 misc_feature
 /note="assembly_name:Contig47"
 misc_feature
 /note="assembly_name:Contig48"
 misc_feature
 /note="assembly_name:Contig49"
 misc_feature
 /note="assembly_name:Contig50
 clone_end:T7
 clone_end:T7
 vector_side:left"
 misc_feature
 /note="assembly_name:Contig51
 clone_end:SP6
 vector_side:right"
 misc_feature
 /note="assembly_name:Contig52"
 misc_feature
 /note="assembly_name:Contig53"

ORIGIN
 Query Match 82.6%; Score 19; DB 2; Length 299377;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCCTCTAAGGAGACCA 19
 Db 177456 GIGCTCTAAGGAGACCA 177438

REFERENCE 1 (bases 1 to 10711)
 AUTHORS Livingston, R.J., Rieder, M.J., Chung, M.-W., Ritchie, T.K., Olson, A.N., Nguyen, C.P., Nguyen, D.A., Poel, C.L., Robertson, P.D., Schackwitz, W.S., Sherwood, J.K., Sherwood, A.M., Leithauser, B.J. and Nickerson, D.A.
 TITLE Direct Submission
 JOURNAL Submitted (12-MAY-2004) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA
 COMMENT To cite this work, please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES5478, Department of Genome Sciences, Seattle, WA (URL: <http://egp.sgs.washington.edu>).
 FEATURES Location_Qualifiers
 1. .100711
 /organism="Homo sapiens"
 /mol_type="Genomic DNA"
 /db_xref="taxon:9606"
 repeat_region 104. .320
 /rpt_family="Alu"
 /rpt_type-dispersed 322. .628
 /rpt_family="Alu"
 /rpt_type-dispersed 611
 variation /frequency="0.01"
 /replace="t"
 repeat_region 639. .770
 /rpt_family="L1"
 /rpt_type-dispersed 867. .978
 /rpt_family="Alu"
 /rpt_type-dispersed 877
 variation /frequency="0.07"
 /replace="g"
 repeat_region 1262. .1506
 /rpt_family="Alu"
 /rpt_type-dispersed 1750
 variation /frequency="0.01"
 /replace="a"
 variation 1947
 /frequency="0.07"
 /replace="c"
 gene 2001. .100192
 /gen="TAF1"
 mRNA 12690. .12931.13343. .13561.13975. .14193.14564. .14771,
 17483. .17659.18276. .18403.18501. .18608.18731. .18904,
 19702. .19875.20685. .20789.23001. .23201.23977. .24118,
 24478. .24608.25325. .25405.28310. .28459.28616. .28735,
 29042. .29217.29808. .29986.32994. .33207.34313. .34478,
 37269. .37480.42379. .42487.43313. .43413.43715. .43892,
 57050. .57117.58858. .59890.59715. .59807.59895. .59979,
 89911. .89978.90482. .90598.93982. .94107.95293. .95449,
 96367. .96544.99566. .100192
 /gene="TAF1"
 /product="TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 250kDa"
 CDS join(2052. .2221,3239. .3353.3794. .3910.10907. .11026,
 12690. .12931.13343. .13561.13975. .14193.14564. .14771,
 17483. .17659.18276. .18403.18501. .18608.18731. .18904,
 19702. .19875.20685. .20789.23001. .23201.23977. .24118,
 24478. .24608.25325. .25405.28310. .28459.28616. .28735,
 29042. .29217.29808. .29986.32994. .33207.34313. .34478,
 37269. .37480.42379. .42487.43313. .43413.43715. .43892,
 57050. .57117.58858. .59890.59715. .59807.59895. .59979,
 89911. .89978.90482. .90598.93982. .94107.95293. .95449,
 96367. .96544.99566. .100192
 /gene="TAF1"
 /codon_start=1
 /product="TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 250kDa"
 /protein_id="AAI38105.1"

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 DEFINITION Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ACCESSION AY623109
 VERSION AY623109.1
 KEYWORDS GI:4777764
 KEYWORDS

/db_xref="GI:4777655" /translation="MGPQCDQEGEVLDDECKRHLAGQALGIGSLLTETDGAUVDKPP
STEADYSDINNEVADEDESRYYQMTGSIQPLCHSDYDDYDAICEDIDCKLNUKPP
PPGPMPKDKDQDSITGVSENGEGIILPSLAPSSLAKEVDFSSSDSESEMGPQEA
TOAESERKCKKCHRELIQEQTOEVCESEVESEVSOKSLWNYDAPPPEQCLSDDEITM
RSARERKCKKCHRELIQEQTOEVCESEVESEVSOKSLWNYDAPPPEQCLSDDEITM
MAPVESXPSQ57GDIOKVDTYDQPKVRYAWRGPUDMUGVPEODSGDGYFKLURKT
EHEPVKSRMIEFREERLEENNGTDLADEFNLMTWHDWDDLIWGEDVYKHKGTKPO
RASLAWHLPSSMTNRMAYNWNQGPAATLDDKPHYSTIEPQDNLNEOLVYGRWEDNIW
AGAMPRLLEPPEVPLTLDENDENLSPKESSESLKGSRILGGT
GVIKEFQOQNSQPEYDQWNLINDEYYPKQGQRLGFTGCGNICHSPAVELROFF
PTPHMGPTKLRFHRPLKCYSGFALSQPGHSVSOPULKHKKKAAMREQDROASGGGE
MFMRITPQDLSLGQDGLLARFYKRPKGDPGAPICKY
FGVPEEBI.KLUSRNLQFQOLQAFENNLFRAPYKHKMPTEFLJRTGQYTRLY
DIEVYQCTSPQPLSLEPGQOLQAFENNLFRAPYKHKMPTEFLJRTGQYTRLY
SHBSSLRKRLLCADEKRIGMSDNNWVLXSPKESSESLKGSRILGGT
AEQRLKDAYGCKSFFPKYVLPKQDKEPOPVKCCTVGTDADELRSLNAKQQLRK
GVADPGCGCGEFSYVLPKQDKEPOPVKCCTVGTDADELRSLNAKQQLRK
FGVPEEBI.KLUSRNLQFQOLQAFENNLFRAPYKHKMPTEFLJRTGQYTRLY
RIFDLOQKVLSLSTEVLISTDSSAEDSFPEMGKNTENMLOKETKSOLSREREQEQ
RKELOMQLLAQGSAAGNNHRDDTASVTSNSASSATGRCLKXKTYTFRDEGKVEYRC
TVRKPAVTDATVTRITYDEEFQKREEMRKEERLQDQLRKLKRQBEQE
KLKGPEPKPKRMKCKPDLKLGAGCAIGMRMTNKFCPLUYQTNAPPSPVANTEEQ
EELEKTVIHDNEELIKVSTKTKVIGKOLJESADEVRSLVLTKEPKQQLPKKKR
DVKLITRMPDQTLRERLDPWVTLSSLBSLINDMDLNTYPTFPTVNAKVK
DVKLITRMPDQTLRERLDPWVTLSSLBSLINDMDLNTYPTFPTVNAKVK
LDCDEKLKEKEKDKLARLKEKAINPLLDDDOQVAPEFILNIVTQRMAYDWSWPFP
VNKFYDYYKVIVNPMDELTQKTSKQYQRSFLDDVNLLANSVYNGPQSY
TKTAQELVNVCQDYLTYDEHDLTQKDCITAKEAALBEAELSDLPMTFGPTPQPP
DLYDENTSLSMRSDASVFODEBNMSMVLDSATAPKQVTOEGEDGDLADEBEGTVO
QPAQASVYEDLMSBEGDDEBAGSDEBEGDNPASQSSGSDVGSGSGIRPKQPR
MLQENTRMDMENEBESMSYEDGEGASHGLQEDNSISGSTYSTEPEDKSNTDTSFSG
GYEVSEEEDEEEEEEQRSGSPVSLQVHLSDEDESDFHSTAGSDLDSDSE"
2198

/gene="TAF1"
/frequency="0.01"
/replace="g"
2587

/gene="TAF1"
/frequency="0.01"
/replace="g"
2620

/gene="TAF1"
/frequency="0.03"
/replace="g"
2841

/gene="TAF1"
/frequency="0.01"
/replace="g"
3005 .3007

/gene="TAF1"
/frequency="0.93"
/replace="n"
3415 .3710

/gene="TAF1"
/frequency="0.93"
/replace="n"
3433

/gene="TAF1"
/frequency="0.03"
/replace="t"
3557

/gene="TAF1"
/frequency="0.02"
/replace="t"
3572

/gene="TAF1"
/frequency="0.03"
/replace="t"
3654

variation 3677 variation 3677

misc_feature 3938 misc_feature 3938

repeat_region 4330 .4571 repeat_region 4330 .4571

repeat_region 4590 .4889 repeat_region 4590 .4889

repeat_region 4890 .507 repeat_region 4890 .507

repeat_region 5342 .562 repeat_region 5342 .562

repeat_region 5684 .5969 repeat_region 5684 .5969

repeat_region 6107 .6389 repeat_region 6107 .6389

repeat_region 7410 .7680 repeat_region 7410 .7680

repeat_region 8505 .8385 repeat_region 8505 .8385

repeat_region 8607 .8805 repeat_region 8607 .8805

repeat_region 8908 .9397 repeat_region 8908 .9397

repeat_region 9682 .9725 repeat_region 9682 .9725

Query Match 81.7% Query Match 81.7%
Best Local Similarity 90.9% Best Local Similarity 90.9%;
Matches 20; Pred. No. 70; Matches 20; Pred. No. 70;
Conservative 0; Mismatches 0; Mismatches 0;
Indels 0; Gaps 0; Gaps 0;

Qy 2 TGTCTAGGAGACACACAGC 23
Db 87026 TGTCTAGGAGACACATCAGC 87047

RESULT 10 AL837520/C AL837520/C
DEFINITION Mouse DNA sequence from clone RP23-41204 on chromosome 2, complete
ACCESSION AL837520
VERSION AL837520.26 GI:32398638
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 12625)
AUTHORS Bates, K.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonesrequest@sanger.ac.uk
COMMENT On Jul 1, 2003 this sequence version replaced gi:31335607.

Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations to either a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate strategy or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL, SW, SWISSPROT; Tr, TRINMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-41204 is from the RPCI-23 Mouse BAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: PBACE:6.

Location/Qualifiers

1..128625

/organism="Mus musculus"

/mol_type="Genomic DNA"

/db_xref="ttxon:10090"

/chromosome="2"

/clone="RP23-41204"

/clone_lib="RPCI-23"

FEATURES

Source

COMMENT

ORIGIN

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

KEYWORDS

VERSION

ORGANISM

DEFINITION

ACCESSION

VERSION

SOURCE

DEFINITION

ACCESSION

VERSION

ORGANISM

DEFINITION

ACCESSION

VERSION

SOURCE

DEFINITION

ACCESSION

VERSION

ORGANISM

DEFINITION

ACCESSION

VERSION

SOURCE

DEFINITION

ACCESSION

VERSION

SOURCE</

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: BMBL; Sw: SWISSPROT; Tr: TRIBMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-474G7 is constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>.

VECTOR: pBAC3.6.
FEATURES
 source
 1. 1.162246
 organism="Mus musculus"
 /mol type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-474G7"
 /clone_id="RPCI-23"
 /clone_end="RPCI-23"

ORIGIN
 Query Match Score 18.8; DB 10; Length 162246;
 Best Local Similarity 90.9%; Pred. No. 68;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

 Qy 2 TGCTCTAAGGAGACCCAGC 23
 Db 95406 TGCTTAAAGAGCACCAACC 95427

RESULT 13
BX572088/c

DEFINITION Mus musculus chromosome X clone RP23-472021, WORKING DRAFT
SEQUENCE, 2 unordered pieces.

VERSION BX572088.2

KEYWORDS HTG, HTGS, PHASE1; HTGS_DRAFT.

SOURCE Mus musculus (house mouse)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 185212)

AUTHORS McLay, K.

TITLE Submitted (31-JUL-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT On Aug 1, 2003 this sequence version replaced gi:33386496.

Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bm472021

----- Summary statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 185051 bases at least Q40
Consensus quality: 185068 bases at least Q30
Consensus quality: 185083 bases at least Q20
Insert size: 185112; sum-of-contigs
Insert size: 183255; 10.0% error; agarose-fp
Quality coverage: 14.20x in Q20 bases; sum-of-contigs Quality coverage: 14.34x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 20274: contig of 20274 bp in length
* 20275 20374: gap of 100 bp
* 20375 185212: contig of 164838 bp in length.
FEATURES
 source
 1. 1.185212
 Location/Qualifiers
 /organism="Mus musculus"
 /mol type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-472021"
 /clone_id="RPCI-23"
 misc_feature
 1. 1.20274
 /notes="assembly fragment:035099
 fragment_chain:1
 clone_end:SP6
 vector_side:left"
 20375. 1.185212
 /notes="assembly fragment:039299
 fragment_chain:1
 clone_end:R7
 vector_side:right"
ORIGIN
 Query Match Score 81.7%; Score 18.8; DB 2; Length 185212;
 Best Local Similarity 90.9%; Pred. No. 67;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 misc_feature
 Qy 2 TGCCTTAAGGAGACCCAGC 23
 Db 78682 TGCCTTAAGGAGACCCAGC 78661
RESULT 14
AC140268
LOCUS AC140268
DEFINITION Mus musculus BAC clone RP23-321K4 from chromosome 7, complete
SEQUENCE
VERSION 1
KEYWORDS HTG,
ORGANISM Mus musculus (house mouse)
REFERENCE 1 (bases 1 to 193572)
AUTHORS Abbott, S.
TITLE The sequence of *Mus musculus* BAC clone RP23-321K4
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 193572)
AUTHORS Wilson, R.
TITLE Sequencing of *Mus musculus*
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 193572)
AUTHORS McPherson, J.D. and Waterston, R.H.

TITLE Direct Submission
JOURNAL Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
REFERENCE Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 193572)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2003) Genome Sequencing Center, 4444 Forest Park
REFERENCE Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 193572)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (25-NOV-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jul 21, 2003 this sequence version replaced gi:28475512.
----- Genome Center
Center Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@watsen.wustl.edu
----- Summary Statistics
Center Project name: M_BA0321k04

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:
 The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
 This sequence is the entire insert of the clone.

FEATURES
source

- 1..193572 "Qualifiers
 - /organism="Mus musculus"
 - /mol_type="genomic DNA"
 - /db_xref="taxon:10050"
 - /chromosome="7"
 - /map="7"
 - /clone_id="RPCI-32K4"
 - /clone.lib="RPCI-23"

repeat_region

- 54..200
 - /rpt_family="Alu"
- 572..628
 - /rpt_family="Alu"
- 657..776
 - /rpt_family="Alu"
- 834..1023
 - /rpt_family="B2"
- 1064..1137
 - /rpt_family="Alu"
- 1138..1164
 - /rpt_family="B4"

repeat_region

- 1294..1574
 - /rpt_family="B4"
- 2448..2622
 - /rpt_family="B2"
- 2625..2754
 - /rpt_family="Alu"
- 2762..2947
 - /rpt_family="B2"
- 3633..3735
 - /rpt_family="Alu"
- 3736..3811
 - /rpt_family="B4"
- 4063..4155
 - /rpt_family="ID"
- 4190..4268
 - /rpt_family="Alu"
- 4293..4410
 - /rpt_family="Li"
- 4411..4507
 - /rpt_family="B2"
- 4611..4756
 - /rpt_family="Alu"
- 4802..4950
 - /rpt_family="Alu"
- 5022..5163
 - /rpt_family="Alu"
- 5235..5343
 - /rpt_family="Alu"
- 5344..5350
 - /rpt_family="Alu"
- 5562..5679
 - /rpt_family="B4"
- 5689..6008
 - /rpt_family="RMERLIB"
- 6297..6527
 - /rpt_family="B2"
- 6620..6813
 - /rpt_family="M4L"
- 7108..7277
 - /rpt_family="M4L"
- 7348..7463
 - /rpt_family="M4L"
- 7529..7734
 - /rpt_family="M4L"
- 8230..8411
 - /rpt_family="M4L"
- 8444..8775
 - /rpt_family="M4L"
- 10057..10145
 - /rpt_family="MIR"
- 10467..10538
 - /rpt_family="B2"
- 10552..10658
 - /rpt_family="B4"
- 12887..13023
 - /rpt_family="Alu"
- 13453..13625
 - /rpt_family="B4"
- 13180..13290
 - /rpt_family="B4"
- 13278..13446
 - /rpt_family="B2"
- 13997..14148
 - /rpt_family="Alu"
- 14393..14553
 - /rpt_family="B4"
- 14597..14759
 - /rpt_family="B2"

tRNA

- complement(14681..14753)
 - /product="tRNA-Ser"
- /note="Likely pseudogene (HMM SC=37.06 / Sec struct

Sc=-11.50)"
repeat_region 15947 .16096 /rpt_family="Alu"
repeat_region 16540 .16752 /rpt_family="B4"
repeat_region 16765 .16911 /rpt_family="Alu"
repeat_region 18092 .18293 /rpt_family="B2"
repeat_region 18308 .18449 /rpt_family="Alu"
repeat_region 18587 .18746 /rpt_family="MAlR"
repeat_region 18641 .18776 /rpt_family="B4"
repeat_region 18818 .19040 /rpt_family="MER2_type"
repeat_region 19630 .19835 /rpt_family="B2"
repeat_region 21195 .21874 /rpt_family="RMR12"
repeat_region 21942 .22091 /rpt_family="Alu"
repeat_region 23102 .23331 /rpt_family="L1"
repeat_region 23286 .23424 /rpt_family="B4"
repeat_region 23610 .23789 /rpt_family="L1"
repeat_region 23794 .23904 /rpt_family="Alu"
repeat_region 24726 .24876 /rpt_family="B2"
repeat_region 25081 .25226 /rpt_family="B4"
repeat_region 25160 .25389 /rpt_family="L1"
repeat_region 25374 .25581 /rpt_family="Alu"
repeat_region 25581 .25726 /rpt_family="B2"
repeat_region 25794 .25931 /rpt_family="B4"
repeat_region 25981 .26178 /rpt_family="L1"
repeat_region 26194 .26391 /rpt_family="Alu"
repeat_region 26406 .26543 /rpt_family="L1"
repeat_region 26619 .26756 /rpt_family="B2"
repeat_region 26831 .26973 /rpt_family="B4"
repeat_region 27044 .27180 /rpt_family="L1"
repeat_region 27257 .27394 /rpt_family="Alu"
repeat_region 27470 .27613 /rpt_family="L1"
repeat_region 27683 .27820 /rpt_family="B2"
repeat_region 27896 .28033 /rpt_family="B4"
repeat_region 28109 .28246 /rpt_family="L1"
repeat_region 28322 .28463 /rpt_family="Alu"
repeat_region 28535 .28672 /rpt_family="L1"
repeat_region 28748 .28885 /rpt_family="B2"
repeat_region 28961 .29098 /rpt_family="B4"
repeat_region 29174 .29315 /rpt_family="L1"
repeat_region 29387 .29524 /rpt_family="Alu"
repeat_region 29599 .29736 /rpt_family="L1"
repeat_region 29812 .29949 /rpt_family="B2"
repeat_region 30025 .30176 /rpt_family="B4"
repeat_region 30238 .30373 /rpt_family="L1"
repeat_region 30451 .30588 /rpt_family="Alu"
repeat_region 30664 .30801 /rpt_family="L1"
repeat_region 30877 .31014 /rpt_family="B2"
repeat_region 31090 .31227 /rpt_family="B4"
repeat_region 31303 .31440 /rpt_family="L1"
repeat_region 31516 .31653 /rpt_family="Alu"
repeat_region 31729 .31866 /rpt_family="L1"
repeat_region 31942 .32079 /rpt_family="B2"
repeat_region 32155 .32292 /rpt_family="B4"
repeat_region 32368 .32505 /rpt_family="L1"
repeat_region 32581 .32718 /rpt_family="Alu"
repeat_region 32794 .32931 /rpt_family="L1"
repeat_region 33007 .33144 /rpt_family="B2"
repeat_region 33220 .33357 /rpt_family="B4"
repeat_region 33433 .33570 /rpt_family="L1"
repeat_region 33646 .33773 /rpt_family="Alu"
repeat_region 33859 .33990 /rpt_family="L1"
repeat_region 34072 .34205 /rpt_family="B2"
repeat_region 34285 .34418 /rpt_family="B4"
repeat_region 34508 .34641 /rpt_family="L1"
repeat_region 34721 .34854 /rpt_family="Alu"
repeat_region 34934 .35067 /rpt_family="L1"
repeat_region 35147 .35280 /rpt_family="B2"
repeat_region 35360 .35493 /rpt_family="B4"
repeat_region 35573 .35706 /rpt_family="L1"
repeat_region 35786 .35919 /rpt_family="Alu"
repeat_region 35999 .36132 /rpt_family="L1"
repeat_region 36212 .36345 /rpt_family="B2"
repeat_region 36425 .36578 /rpt_family="B4"
repeat_region 36638 .36771 /rpt_family="L1"
repeat_region 36851 .36984 /rpt_family="Alu"
repeat_region 37064 .37197 /rpt_family="L1"
repeat_region 37277 .37410 /rpt_family="B2"
repeat_region 37490 .37623 /rpt_family="B4"
repeat_region 37703 .37836 /rpt_family="L1"
repeat_region 37916 .38049 /rpt_family="Alu"
repeat_region 38129 .38262 /rpt_family="L1"
repeat_region 38342 .38475 /rpt_family="B2"
repeat_region 38555 .38688 /rpt_family="B4"
repeat_region 38768 .38901 /rpt_family="L1"
repeat_region 39081 .39214 /rpt_family="Alu"
repeat_region 39294 .39427 /rpt_family="L1"
repeat_region 39507 .39640 /rpt_family="B2"
repeat_region 39720 .39853 /rpt_family="B4"
repeat_region 39933 .40066 /rpt_family="L1"
repeat_region 40146 .40279 /rpt_family="Alu"
repeat_region 40359 .40492 /rpt_family="L1"
repeat_region 40572 .40705 /rpt_family="B2"
repeat_region 40785 .40918 /rpt_family="B4"
repeat_region 41098 .41231 /rpt_family="L1"
repeat_region 41311 .41444 /rpt_family="Alu"
repeat_region 41524 .41657 /rpt_family="L1"
repeat_region 41737 .41870 /rpt_family="B2"
repeat_region 41950 .42083 /rpt_family="B4"
repeat_region 42163 .42296 /rpt_family="L1"
repeat_region 42376 .42509 /rpt_family="Alu"
repeat_region 42589 .42722 /rpt_family="L1"
repeat_region 42702 .42835 /rpt_family="B2"
repeat_region 42915 .43048 /rpt_family="B4"
repeat_region 43128 .43261 /rpt_family="L1"
repeat_region 43341 .43474 /rpt_family="Alu"
repeat_region 43554 .43687 /rpt_family="L1"
repeat_region 43767 .43890 /rpt_family="B2"
repeat_region 43980 .44113 /rpt_family="B4"
repeat_region 44193 .44326 /rpt_family="L1"
repeat_region 44406 .44539 /rpt_family="Alu"
repeat_region 44619 .44752 /rpt_family="L1"
repeat_region 44832 .44965 /rpt_family="B2"
repeat_region 45045 .45178 /rpt_family="B4"
repeat_region 45258 .45391 /rpt_family="L1"
repeat_region 45471 .45584 /rpt_family="Alu"
repeat_region 45684 .45817 /rpt_family="L1"
repeat_region 45897 .46020 /rpt_family="B2"
repeat_region 46110 .46243 /rpt_family="B4"
repeat_region 46323 .46456 /rpt_family="L1"
repeat_region 46536 .46669 /rpt_family="Alu"
repeat_region 46749 .46882 /rpt_family="L1"
repeat_region 46962 .47095 /rpt_family="B2"
repeat_region 47175 .47308 /rpt_family="B4"
repeat_region 47388 .47521 /rpt_family="L1"
repeat_region 47501 .47634 /rpt_family="Alu"
repeat_region 47714 .47847 /rpt_family="L1"
repeat_region 47927 .48060 /rpt_family="B2"
repeat_region 48140 .48273 /rpt_family="B4"
repeat_region 48353 .48486 /rpt_family="L1"
repeat_region 48566 .48699 /rpt_family="Alu"
repeat_region 48779 .48912 /rpt_family="L1"
repeat_region 49002 .49135 /rpt_family="B2"
repeat_region 49215 .49348 /rpt_family="B4"
repeat_region 49428 .49561 /rpt_family="L1"
repeat_region 49641 .49774 /rpt_family="Alu"
repeat_region 49854 .49987 /rpt_family="L1"
repeat_region 50067 .50190 /rpt_family="B2"
repeat_region 50280 .50413 /rpt_family="B4"
repeat_region 50493 .50626 /rpt_family="L1"
repeat_region 50706 .50839 /rpt_family="Alu"
repeat_region 50919 .51052 /rpt_family="L1"
repeat_region 51132 .51265 /rpt_family="B2"
repeat_region 51345 .51478 /rpt_family="B4"
repeat_region 51558 .51691 /rpt_family="L1"
repeat_region 51771 .51894 /rpt_family="Alu"
repeat_region 51984 .52117 /rpt_family="L1"
repeat_region 52197 .52330 /rpt_family="B2"
repeat_region 52410 .52543 /rpt_family="B4"
repeat_region 52623 .52756 /rpt_family="L1"
repeat_region 52836 .52969 /rpt_family="Alu"
repeat_region 53049 .53182 /rpt_family="L1"
repeat_region 53262 .53395 /rpt_family="B2"
repeat_region 53475 .53608 /rpt_family="B4"
repeat_region 53688 .53811 /rpt_family="L1"
repeat_region 53901 .54034 /rpt_family="Alu"
repeat_region 54114 .54237 /rpt_family="L1"
repeat_region 54327 .54460 /rpt_family="B2"
repeat_region 54540 .54673 /rpt_family="B4"
repeat_region 54753 .54886 /rpt_family="L1"
repeat_region 54966 .55099 /rpt_family="Alu"
repeat_region 55179 .55302 /rpt_family="L1"
repeat_region 55392 .55525 /rpt_family="B2"
repeat_region 55605 .55738 /rpt_family="B4"
repeat_region 55818 .55951 /rpt_family="L1"
repeat_region 56031 .56164 /rpt_family="Alu"
repeat_region 56244 .56377 /rpt_family="L1"
repeat_region 56457 .56590 /rpt_family="B2"
repeat_region 56670 .56783 /rpt_family="B4"
repeat_region 56883 .57006 /rpt_family="L1"
repeat_region 57096 .57229 /rpt_family="Alu"
repeat_region 57309 .57442 /rpt_family="L1"
repeat_region 57522 .57655 /rpt_family="B2"
repeat_region 57735 .57868 /rpt_family="B4"
repeat_region 57948 .58071 /rpt_family="L1"
repeat_region 58161 .58294 /rpt_family="Alu"
repeat_region 58374 .58507 /rpt_family="L1"
repeat_region 58587 .58720 /rpt_family="B2"
repeat_region 58800 .58913 /rpt_family="B4"
repeat_region 59013 .59136 /rpt_family="L1"
repeat_region 59226 .59359 /rpt_family="Alu"
repeat_region 59439 .59572 /rpt_family="L1"
repeat_region 59652 .59775 /rpt_family="B2"
repeat_region 59865 .59998 /rpt_family="B4"
repeat_region 60078 .60201 /rpt_family="L1"
repeat_region 60291 .60414 /rpt_family="Alu"
repeat_region 60504 .60627 /rpt_family="L1"
repeat_region 60717 .60840 /rpt_family="B2"
repeat_region 60930 .61053 /rpt_family="B4"
repeat_region 61143 .61266 /rpt_family="L1"
repeat_region 61356 .61479 /rpt_family="Alu"
repeat_region 61569 .61692 /rpt_family="L1"
repeat_region 61782 .61905 /rpt_family="B2"
repeat_region 61995 .62118 /rpt_family="B4"
repeat_region 62208 .62331 /rpt_family="L1"
repeat_region 62421 .62544 /rpt_family="Alu"
repeat_region 62634 .62757 /rpt_family="L1"
repeat_region 62847 .62970 /rpt_family="B2"
repeat_region 63060 .63183 /rpt_family="B4"
repeat_region 63273 .63396 /rpt_family="L1"
repeat_region 63486 .63609 /rpt_family="Alu"
repeat_region 63700 .63823 /rpt_family="L1"
repeat_region 63913 .64036 /rpt_family="B2"
repeat_region 64126 .64249 /rpt_family="B4"
repeat_region 64339 .64462 /rpt_family="L1"
repeat_region 64552 .64675 /rpt_family="Alu"
repeat_region 64765 .64888 /rpt_family="L1"
repeat_region 64978 .65101 /rpt_family="B2"
repeat_region 65191 .65314 /rpt_family="B4"
repeat_region 65404 .65527 /rpt_family="L1"
repeat_region 65617 .65740 /rpt_family="Alu"
repeat_region 65830 .65953 /rpt_family="L1"
repeat_region 66043 .66166 /rpt_family="B2"
repeat_region 66256 .66379 /rpt_family="B4"
repeat_region 66469 .66592 /rpt_family="L1"
repeat_region 66682 .66805 /rpt_family="Alu"
repeat_region 66895 .67018 /rpt_family="L1"
repeat_region 67108 .67231 /rpt_family="B2"
repeat_region 67321 .67444 /rpt_family="B4"
repeat_region 67534 .67657 /rpt_family="L1"
repeat_region 67747 .67870 /rpt_family="Alu"
repeat_region 67960 .68083 /rpt_family="L1"
repeat_region 68173 .68296 /rpt_family="B2"
repeat_region 68386 .68509 /rpt_family="B4"
repeat_region 68599 .68722 /rpt_family="L1"
repeat_region 68812 .68935 /rpt_family="Alu"
repeat_region 69025 .69148 /rpt_family="L1"
repeat_region 69238 .69361 /rpt_family="B2"
repeat_region 69451 .69574 /rpt_family="B4"
repeat_region 69664 .69787 /rpt_family="L1"
repeat_region 69877 .69990 /rpt_family="Alu"
repeat_region 70090 .70213 /rpt_family="L1"
repeat_region 70303 .70426 /rpt_family="B2"
repeat_region 70516 .70639 /rpt_family="B4"
repeat_region 70729 .70852 /rpt_family="L1"
repeat_region 70942 .71065 /rpt_family="Alu"
repeat_region 71155 .71278 /rpt_family="L1"
repeat_region 71368 .71491 /rpt_family="B2"
repeat_region 71581 .71704 /rpt_family="B4"
repeat_region 71794 .71917 /rpt_family="L1"
repeat_region 72007 .72130 /rpt_family="Alu"
repeat_region 72220 .72343 /rpt_family="L1"
repeat_region 72433 .72556 /rpt_family="B2"
repeat_region 72646 .72769 /rpt_family="B4"
repeat_region 72859 .72982 /rpt_family="L1"
repeat_region 73072 .73195 /rpt_family="Alu"
repeat_region 73285 .73408 /rpt_family="L1"
repeat_region 73498 .73621 /rpt_family="B2"
repeat_region 73711 .73834 /rpt_family="B4"
repeat_region 73924 .74047 /rpt_family="L1"
repeat_region 74137 .74260 /rpt_family="Alu"
repeat_region 74350 .74473 /rpt_family="L1"
repeat_region 74563 .74686 /rpt_family="B2"
repeat_region 74776 .74899 /rpt_family="B4"
repeat_region 74989 .75112 /rpt_family="L1"
repeat_region 75202 .75325 /rpt_family="Alu"
repeat_region 75415 .75538 /rpt_family="L1"
repeat_region 75628 .75751 /rpt_family="B2"
repeat_region 75841 .75964 /rpt_family="B4"
repeat_region 76054 .76177 /rpt_family="L1"
repeat_region 76267 .76390 /rpt_family="Alu"
repeat_region 76480 .76593 /rpt_family="L1"
repeat_region 76693 .76816 /rpt_family="B2"
repeat_region 76906 .77029 /rpt_family="B4"
repeat_region 77119 .77242 /rpt_family="L1"
repeat_region 77332 .77455 /rpt_family="Alu"
repeat_region 77545 .77668 /rpt_family="L1"
repeat_region 77758 .77881 /rpt_family="B2"
repeat_region 77971 .78094 /rpt_family="B4"
repeat_region 78184 .78307 /rpt_family="L1"
repeat_region 78397 .78520 /rpt_family="Alu"
repeat_region 78610 .78733 /rpt_family="L1"
repeat_region 78823 .78946 /rpt_family="B2"
repeat_region 79036 .79159 /rpt_family="B4"
repeat_region 79249 .79372 /rpt_family="L1"
repeat_region 79462 .79585 /rpt_family="Alu"
repeat_region 79675 .79798 /rpt_family="L1"
repeat_region 79888 .79911 /rpt_family="B2"
repeat_region 80101 .80224 /rpt_family="B4"
repeat_region 80314 .80437 /rpt_family="L1"
repeat_region 80527 .80650 /rpt_family="Alu"
repeat_region 80740 .80863 /rpt_family="L1"
repeat_region 80953 .81076 /rpt_family="B2"
repeat_region 81166 .81289 /rpt_family="B4"
repeat_region 81379 .81492 /rpt_family="L1"
repeat_region 81592 .81715 /rpt_family="Alu"
repeat_region 81805 .81928 /rpt_family="L1"
repeat_region 82018 .82141 /rpt_family="B2"
repeat_region 82231 .82354 /rpt_family="B4"
repeat_region 82444 .82567 /rpt_family="L1"
repeat_region 82657 .82780 /rpt_family="Alu"
repeat_region 82870 .82993 /rpt_family="L1"
repeat_region 83083 .83206 /rpt_family="B2"
repeat_region 83296 .83419 /rpt_family="B4"
repeat_region 83509 .83622 /rpt_family="L1"
repeat_region 83722 .83845 /rpt_family="Alu"
repeat_region 83935 .84048 /rpt_family="L1"
repeat_region 84148 .84271 /rpt_family="B2"
repeat_region 84361 .84484 /rpt_family="B4"
repeat_region 84574 .84697 /rpt_family="L1"
repeat_region 84787 .84910 /rpt_family="Alu"
repeat_region 85000 .85123 /rpt_family="L1"
repeat_region 85213 .85336 /rpt_family="B2"
repeat_region 85426 .85549 /rpt_family="B4"
repeat_region 85639 .85752 /rpt_family="L1"
repeat_region 85852 .85975 /rpt_family="Alu"
repeat_region 86065 .86188 /rpt_family="L1"
repeat_region 86278 .86401 /rpt_family="B2"
repeat_region 86491 .86614 /rpt_family="B4"
repeat_region 86704 .86827 /rpt_family="L1"
repeat_region 86917 .87040 /rpt_family="Alu"
repeat_region 87130 .87253 /rpt_family="L1"
repeat_region 87343 .87466 /rpt_family="B2"
repeat_region 87556 .87669 /rpt_family="B4"
repeat_region 87769 .87892 /rpt_family="L1"
repeat_region 87982 .88105 /rpt_family="Alu"
repeat_region 88195 .88318 /rpt_family="L1"
repeat_region 88408 .88531 /rpt_family="B2"
repeat_region 88621 .88734 /rpt_family="B4"
repeat_region 88834 .88947 /rpt_family="L1"
repeat_region 89047 .89170 /rpt_family="Alu"
repeat_region 89260 .89373 /rpt_family="L1"
repeat_region 89473 .89586 /rpt_family="B2"
repeat_region 89686 .89809 /rpt_family="B4"
repeat_region 89900 .90023 /rpt_family="L1"
repeat_region 90113 .90236 /rpt_family="Alu"
repeat_region 90326 .90439 /rpt_family="L1"
repeat_region 90539 .90652 /rpt_family="B2"
repeat_region 90752 .90865 /rpt_family="B4"
repeat_region 90965 .91078 /rpt_family="L1"
repeat_region 91178 .91291 /rpt_family="Alu"
repeat_region 91391 .91504 /rpt_family="L1"
repeat_region 91604 .91717 /rpt_family="B2"
repeat_region 91817 .91930 /rpt_family="B4"
repeat_region 92030 .92143 /rpt_family="L1"
repeat_region 92243 .92356 /rpt_family="Alu"
repeat_region 92456 .92569 /rpt_family="L1"
repeat_region 92669 .92782 /rpt_family="B2"
repeat_region 92882 .93095 /rpt_family="B4"
repeat_region 93105 .93218 /rpt_family="L1"
repeat_region 93318 .93431 /rpt_family="Alu"
repeat_region 93531 .93644 /rpt_family="L1"
repeat_region 93744 .93857 /rpt_family="B2"
repeat_region 93957 .94070 /rpt_family="B4"
repeat_region 94170 .94283 /rpt_family="L1"
repeat_region 94383 .94496 /rpt_family="Alu"
repeat_region 94596 .94709 /rpt_family="L1"
repeat_region 94809 .94922 /rpt_family="B2"
repeat_region 95022 .95135 /rpt_family="B4"
repeat_region 95235 .95348 /rpt_family="L1"
repeat_region 95448 .95561 /rpt_family="Alu"
repeat_region 95661 .95774 /rpt_family="L1"
repeat_region 95874 .95987 /rpt_family="B2"
repeat_region 96087 .96200 /rpt_family="B4"
repeat_region 96300 .96413 /rpt_family="L1"
repeat_region 96513 .96626 /rpt_family="Alu"
repeat_region 96726 .96839 /rpt_family="L1"
repeat_region 96939 .97052 /rpt_family="B2"
repeat_region 97152 .97265 /rpt_family="B4"
repeat_region 97365 .97478 /rpt_family="L1"
repeat_region 97578 .97691 /rpt_family="Alu"
repeat_region 97791 .97904 /rpt_family="L1"
repeat_region 98004 .98117 /rpt_family="B2"
repeat_region 98217 .98330 /rpt_family="B4"
repeat_region 98430 .98543 /rpt_family="L1"
repeat_region 98643 .98756 /rpt_family="Alu"
repeat_region 98856 .98969 /rpt_family="L1"
repeat_region 99069 .99182 /rpt_family="B2"
repeat_region 99282 .99495 /rpt_family="B4"
repeat_region 99505 .99618 /rpt_family="L1"
repeat_region 99718 .99831 /rpt_family="Alu"
repeat_region 99931 .99944 /rpt_family="L1"
repeat_region 100144 .100257 /rpt_family="B2"
repeat_region 100357 .100470 /rpt_family="B4"
repeat_region 100570 .100683 /rpt_family="L1"
repeat_region 100783 .100896 /rpt_family="Alu"
repeat_region 100996 .101109 /rpt_family="L1"
repeat_region 101209 .101322 /rpt_family="B2"
repeat_region 101422 .101535 /rpt_family="B4"
repeat_region 101635 .101748 /rpt_family="L1"
repeat_region 101848 .101961 /rpt_family="Alu"
repeat_region 102061 .102174 /rpt_family="L1"
repeat_region 102274 .102387 /rpt_family="B2"
repeat_region 102487 .102500 /rpt_family="B4"
repeat_region 102690 .102703 /rpt_family="L1"
repeat_region 102803 .102916 /rpt_family="Alu"
repeat_region 103016 .103129 /rpt_family="L1"
repeat_region 103229 .103342 /rpt_family="B2"
repeat_region 103442 .103555 /rpt_family="B4"
repeat_region 103655 .103768 /rpt_family="L1"
repeat_region 103868 .103981 /rpt_family="Alu"
repeat_region 104081 .104194 /rpt_family="L1"
repeat_region 104294 .104407 /rpt_family="B2"
repeat_region 104507 .104620 /rpt_family="B4"
repeat_region 104720 .104833 /rpt_family="L1"
repeat_region 104933 .105046 /rpt_family="Alu"
repeat_region 105146 .105259 /rpt_family="L1"
repeat_region 105359 .105472 /rpt_family="B2"
repeat_region 105572 .105685 /rpt_family="B4"
repeat_region 105785 .105898 /rpt_family="L1"
repeat_region 105998 .106111 /rpt_family="Alu"
repeat_region 106211 .106324 /rpt_family="L1"
repeat_region 106424 .106537 /rpt_family="B2"
repeat_region 106637 .106750 /rpt_family="B4"
repeat_region 106850 .106963 /rpt_family="L1"
repeat_region 107063 .107176 /rpt_family="Alu"
repeat_region 107276 .107389 /rpt_family="L1"
repeat_region 107489 .107502 /rpt_family="B2"
repeat_region 107602 .107715 /rpt_family="B4"
repeat_region 107815 .107928 /rpt_family="L1"
repeat_region 108028 .108141 /rpt_family="Alu"
repeat_region 108241 .108354 /rpt_family="L1"
repeat_region 108454 .108567 /rpt_family="B2"
repeat_region 108667 .108780 /rpt_family="B4"
repeat_region 108880 .108993 /rpt_family="L1"
repeat_region 109093 .109206 /rpt_family="Alu"
repeat_region 109306 .109419 /rpt_family="L1"
repeat_region 109519 .109632 /rpt_family="B2"
repeat_region 109732 .109845 /rpt_family="B4"
repeat_region 109945 .110058 /rpt_family="L1"
repeat_region 110158 .110271 /rpt_family="Alu"
repeat_region 110371 .110484 /rpt_family="L1"
repeat_region 110584 .110697 /rpt_family="B2"
repeat_region 110797 .110910 /rpt_family="B4"
repeat_region 111010 .111123 /rpt_family="L1"
repeat_region 111223 .111336 /rpt_family="Alu"
repeat_region 111436 .111549 /rpt_family="L1"
repeat_region 111649 .111762 /rpt_family="B2"
repeat_region 111862 .111975 /rpt_family="B4"
repeat_region 112075 .112188 /rpt_family="L1"
repeat_region 112288 .112401 /rpt_family="Alu"
repeat_region 112401 .112514 /rpt_family="L1"
repeat_region 112614 .112727 /rpt_family="B2"
repeat_region 112827 .112940 /rpt_family="B4"
repeat_region 113040 .113153 /rpt_family="L1"
repeat_region 113253 .113366 /rpt_family="Alu"
repeat_region 113466 .113579 /rpt_family="L1"
repeat_region 113679 .113792 /rpt_family="B2"
repeat_region 113892 .114005 /rpt_family="B4"
repeat_region 114105 .114218 /rpt_family="L1"
repeat_region 114318 .114431 /rpt_family="Alu"
repeat_region 114531 .114644 /rpt_family="L1"
repeat_region 114744 .114857 /rpt_family="B2"
repeat_region 114957 .115070 /rpt_family="B4"
repeat_region 115170 .115283 /rpt_family="L1"
repeat_region 115383 .115496 /rpt_family="Alu"
repeat_region 115596 .115709 /rpt_family="L1"
repeat_region 115809 .115922 /rpt_family="B2"
repeat_region 116022 .116135 /rpt_family="B4"
repeat_region 116235 .116348 /rpt_family="L1"
repeat_region 116448 .116561 /rpt_family="Alu"
repeat_region 116661 .116774 /rpt_family="L1"
repeat_region 116874 .116987 /rpt_family="B2"
repeat_region 117087 .117200 /rpt_family="B4"
repeat_region 117300 .117413 /rpt_family="L1"
repeat_region 117513 .117626 /rpt_family="Alu"
repeat_region 117726 .117839 /rpt_family="L1"
repeat_region 117939 .118052 /rpt_family="B2"
repeat_region 118152 .118265 /rpt_family="B4"
repeat_region 118365 .118478 /rpt_family="L1"
repeat_region 118578 .118691 /rpt_family="Alu"
repeat_region 118791 .118904 /rpt_family="L1"
repeat_region 119004 .119117 /rpt_family="B2"
repeat

GenCore version 5.1.6
 Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: March 9, 2005, 02:14:38 ; Search time 427 Seconds
 (without alignments)
 318,862 Million cell updates/sec

Title: US-09-835-976B-115

Perfect score: 23

Sequence: 1 gtgtcttaaggagacaccacagg 23

Scoring table: IDENTITY_NUC
 Gapoff 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing First 45 summaries

Database : N_Geneseq_16Dec04:*

1: geneseqn1980s:*

2: geneseqn1930s:*

3: geneseqn2000s:*

4: geneseqn2001s:*

5: geneseqn2001bs:*

6: geneseqn2000a:*

7: geneseqn2002bs:*

8: geneseqn2003a:*

9: geneseqn2003bs:*

10: geneseqn2003cs:*

11: geneseqn2003ds:*

12: geneseqn2004a:*

13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description	
1	23	100.0	23	4	AA59492	AA59492 Human KCC	
2	18.2	79.1	231004	12	AD097855	Adg97855 Mouse can	
3	18	78.3	487	6	ABK71639	Abk71639 Human dit	
4	18	78.3	1287	12	AD000741	Ad000741 Novel hum	
5	18	78.3	1287	12	AD999172	Ad999172 Novel hum	
6	18	78.3	1726	5	AAST2172	Aas72172 DNA encod	
7	18	78.3	1766	12	AChB7437	Achb7437 Human gen	
8	18	78.3	2077	3	Aaz52430	Aaz52430 HTMR clon	
c	9	17.8	77.4	1646	8	ABX15830	Abx15830 cDNA enco
c	10	17.8	77.4	1645	12	AD118998	Adi118998 Soybean s
c	11	17.8	77.4	52302	9	ADA02738	Ada02738 Human CCN
c	12	17.8	77.4	52302	10	ADB72476	Adb72476 Human CCN
c	13	17.8	77.4	52302	10	ADC85218	Adc85218 Human Ccn
c	14	17.8	77.4	52302	12	ADM74333	Adm74333 Human car
c	15	17.8	77.4	14405	11	ACN44062	Acn44062 Human gen
c	16	17.2	74.8	146	12	ADK43275	Adk43275 Murine pr
c	17	17.2	74.8	761	10	ACD94539	Acd94539 Human col
c	18	17.2	74.8	1872	3	AAC47169	Aaa47169 DNA seqne
c	19	17.2	74.8	2075	2	AAV16467	Aav16467 Nucleotid
c	20	17.2	74.8	2091	2	AAX18839	Streptoco

ALIGNMENTS

RESULT 1
 AA59492
 ID AA59492 standard; DNA; 23 BP.
 XX
 XX
 AC
 XX
 DT 16-JAN-2002 (first entry)
 XX
 DE Human KCC2 polymorphism detection primer #2.
 XX
 KW Human; ss; potassium-chloride cotransporter; epilepsy; hypertension; KCC;
 KW hypotensive; anticonvulsant; antianemic; sickle-cell anaemia;
 KW gene therapy; PCR primer.
 XX
 OS Homo sapiens.
 XX
 PN WO200179525-A2.
 XX
 XX
 XX
 XX
 XX
 XX
 PR 14-APR-2000; 2000US-0197350P.
 XX
 (UYVA-) UNIV VANDERBILT.
 PA (GAMB/) GAMBA G.
 XX
 PI Gamba G, Mount DB, Delpire E, George AL;
 XX
 DR WPI; 2001-611726/70.
 XX
 PT New isolated and purified potassium-chloride cotransporter polypeptide
 PR for detecting a modulator of potassium-chloride cotransport that can
 PR treat epilepsy, sickle cell anemia, and hypertension.
 XX
 PS Example 26; Page 347; 352pp; English.
 XX
 CC The invention relates to an isolated and purified, biologically active
 CC KCC2, KCC3, or KCC4 potassium-chloride cotransporter polypeptide and
 CC polynucleotides encoding them. KCC and the nucleic acid encoding KCC are
 CC used to produce an antibody against KCC, which can be used to detect KCC.
 CC The antibody and nucleic acid are used in kits to detect KCC and the
 CC nucleic acid encoding KCC. KCC and nucleic acid encoding it are used to

CC screen for candidate substances that can modulate the activity of KCC.
 CC KCC and the nucleic acid encoding it can be used to modulate potassium-
 CC chloride cotransport in a vertebrate. Therapeutic compositions comprising
 CC modulators of the activity of KCC (preferably antibodies or antisense
 CC oligonucleotides) are used to treat epilepsy, sickle cell anemia, and
 CC hypertension (also via gene therapy using the KCC polynucleotide). The
 CC present sequence is a PCR primer used to detect polymorphisms in a DNA
 XX sequence encoding a KCC of the invention

SQ Sequence 23 BP; 7 A; 6 C; 6 G; 4 T; 0 U; 0 Other;

Query Match 100.0%; Score 23; DB 4; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.36;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGCTTAAGGAGACCCAGC 23

Db 1 GTGCTTAAGGAGACCCAGC 23

ABK71639
 ID ABK71639 standard; cDNA; 487 BP.

XX XX ABK71639;

AC XX 30-JUL-2002 (first entry)

DT XX Human dithp polynucleotide #105.

DE XX Human; dithp; diagnostic and therapeutic polynucleotide; gene; ss; bone;

KW cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;

KW inflammatory disorder; viral infection; bacterial infection; seizure;

KW fungal infection; parasitic infections; developmental disorder; breast;

KW endocrine disorder; metabolic disorder; neurological disorder; cervix;

KW gastrointestinal disorder; transport disorder; gene therapy; kidney;

KW adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;

KW skin; testis; thymus.

XX XX Homo sapiens.

OS XX WO200220754-A2.

PN XX WO200220754.

PD XX 14-MAR-2002.

PP XX 29-AUG-2001; 2001WO-US027127.

XX XX PR 05-SEP-2000; 2000US-0229747P.

PR 05-SEP-2000; 2000US-0229748P.

PR 05-SEP-2000; 2000US-0229749P.

PR 05-SEP-2000; 2000US-0229750P.

PR 05-SEP-2000; 2000US-0229751P.

PR 05-SEP-2000; 2000US-0230533P.

PR 06-SEP-2000; 2000US-0230505P.

PR 06-SEP-2000; 2000US-0230514P.

PR 06-SEP-2000; 2000US-0230515P.

PR 06-SEP-2000; 2000US-0230517P.

PR 06-SEP-2000; 2000US-0230518P.

PR 06-SEP-2000; 2000US-0230519P.

PR 06-SEP-2000; 2000US-02305195P.

PR 06-SEP-2000; 2000US-0230597P.

PR 06-SEP-2000; 2000US-0230598P.

PR 06-SEP-2000; 2000US-0230599P.

PR 06-SEP-2000; 2000US-0230610P.

PR 06-SEP-2000; 2000US-0230685P.

PR 06-SEP-2000; 2000US-0230988P.

PR 07-SEP-2000; 2000US-0230931P.

PR 07-SEP-2000; 2000US-0231163P.

PR 07-SEP-2000; 2000US-0231167P.

XX XX (INCYT-) INCYTE GENOMICS INC.

PA PA

XX XX P1 Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;

P1 Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TP, Yap PE, Dahl CR;

P1 Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Rosberry AM;

P1 Gerstlin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;

P1 Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;

XX XX DR WPI; 2002-383054/41.

XX XX P-PSDB; AB66048.

PT An isolated polynucleotide useful in diagnostics and therapeutics.

XX XX Claim 1; Page 454; 686pp; English.

XX XX The invention relates to human diagnostic and therapeutic (dithp) polypeptides (DIRNP polypeptides).

CC The sequences of the invention are used in the treatment and diagnosis of

CC cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers

CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,

CC cervix, kidney, lung, ovary, pancreas, prostate, spleen, bronchitis or

CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,

CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal

CC infections, parasitic infections, developmental disorders (e.g. anaemia,

CC RESULT 3

CC epilepsy), seizure disorders (e.g. cerebral palsy, spina bifida),
 CC endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders
 CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,
 CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal
 CC disorders (e.g. ulcerative colitis, lymphinuria) and transport disorders
 CC (e.g. myotonic dystrophy, catatonia, peripheral neuropathy). Sequences
 CC ABK71535-ABK71809 represent human dithp polynucleotides of the invention
 XX Sequence 487 BP, 162 A; 94 C; 101 G; 130 T; 0 U; 0 Other;
 SQ PR 03-JAN-2003; 2003US-0437914P.
 SQ PR 17-JAN-2003; 2003US-0440220P.
 CC PR 17-JAN-2003; 2003US-0440221P.
 CC PR 18-APR-2003; 2003US-0463700P.
 CC PR 18-APR-2003; 2003US-0463708P.
 CC PR 18-APR-2003; 2003US-0463716P.
 CC PR 18-APR-2003; 2003US-0463732P.
 CC PR 02-MAY-2003; 2003US-0467199P.
 CC PR 02-MAY-2003; 2003US-0467201P.
 CC PR 02-MAY-2003; 2003US-0467203P.
 CC PR 02-MAY-2003; 2003US-0467230P.
 CC PR 19-MAY-2003; 2003US-0471206P.
 CC PR 19-MAY-2003; 2003US-047136P.
 CC PR 22-MAY-2003; 2003US-0472420P.
 CC PR 22-MAY-2003; 2003US-0472430P.
 CC PR 09-JUN-2003; 2003US-0476509P.
 CC PR 09-JUN-2003; 2003US-0476621P.
 CC PR 09-JUN-2003; 2003US-0476532P.
 CC PR 09-JUN-2003; 2003US-0476541P.
 CC PR 08-JUL-2003; 2003US-0472417P.
 CC PR 08-JUL-2003; 2003US-0485201P.
 CC PR 08-JUL-2003; 2003US-048523P.
 CC PR 08-JUL-2003; 2003US-0485224P.
 CC PR 08-JUL-2003; 2003US-048525P.
 CC PR 08-JUL-2003; 2003US-048559P.
 CC PR 14-JUL-2003; 2003US-0485446P.
 CC PR 14-JUL-2003; 2003US-0485480P.
 CC PR 15-JUL-2003; 2003US-0486591P.
 CC PR 15-JUL-2003; 2003US-048660P.
 CC PR 08-AUG-2003; 2003US-0493341P.
 CC PR 08-AUG-2003; 2003US-0493370P.
 CC PR 08-AUG-2003; 2003US-0493373P.
 CC PR 08-AUG-2003; 2003US-0493577P.
 XX PA (FIVE) FIVE PRIME THERAPEUTICS INC.
 XX XX
 PI Williams LT, Chu K, Lee B, Hestir K, Beauroang PA, Behrens D;
 PI Halenbeck RF, Kothakota S, Lin H, Lirnemann T, Pierce K, Wang Y;
 PI Wong JGP, Wu G, Zhang H, Zeng C;
 XX XX
 DR WPI; 2004-365511/34.
 DR P-PSDB; ADN99956.

RESULT 4
 ADO00741 standard; cDNA; 1287 BP.
 XX
 ADO00741;
 XX
 29-JUL-2004 (First entry)
 XX Novel human cDNA sequence #1556.
 DB ds; gene; anti-inflammatory; dermatological; neuroprotective;
 DB immunomodulator; antibacterial; virucide; antipsoriatic;
 DB cryostatic;
 KW gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;
 KW psoriasis; diabetes; early aging; hormonal imbalance;
 KW ischemic heart disease; ulcerative colitis.
 KW XX
 OS Homo sapiens.
 XX PA
 XX WO2004038003-A2.
 XX PN
 XX PD 06-MAY-2004.
 XX PP 24-OCT-2003; 2003WO-US033947.
 XX PR 25-OCT-2002; 2002US-0421061P.
 PR 25-OCT-2002; 2002US-0421080P.
 PR 25-OCT-2002; 2002US-0421557P.
 PR 25-OCT-2002; 2002US-0421614P.
 PR 30-OCT-2002; 2002US-0421777P.
 PR 30-OCT-2002; 2002US-0422178P.
 PR 15-NOV-2002; 2002US-0426355P.
 PR 15-NOV-2002; 2002US-0426384P.
 PR 27-NOV-2002; 2002US-0426394P.
 PR 15-NOV-2002; 2002US-0426410P.
 PR 15-NOV-2002; 2002US-0426451P.
 PR 15-NOV-2002; 2002US-042916P.
 PR 27-NOV-2002; 2002US-0429224P.
 PR 27-NOV-2002; 2002US-0429275P.
 PR 27-NOV-2002; 2002US-0429302P.
 PR 27-NOV-2002; 2002US-0429326P.
 PR 04-DEC-2002; 2002US-0430663P.
 PR 04-DEC-2002; 2002US-0430665P.
 PR 04-DEC-2002; 2002US-0430677P.
 PR 04-DEC-2002; 2002US-0430683P.
 PR 04-DEC-2002; 2002US-0431458P.
 PR 12-DEC-2002; 2002US-043251P.
 PR 04-DEC-2002; 2002US-0430649P.
 PR 05-DEC-2002; 2002US-0430937P.
 PR 05-DEC-2002; 2002US-0430965P.
 PR 12-DEC-2002; 2002US-0433500P.
 PR 13-DEC-2002; 2002US-0433316P.
 PR 13-DEC-2002; 2002US-0433318P.
 PR 23-DEC-2002; 2002US-0436236P.

Query Match 78.3%; Score 18; DB 12; Length 1287;
 Best Local Similarity 100.0%; Pred. No. 1.5e-02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

4 TCTTAAGGAGACCAACA 21
 59.TCTTAAGGAGACCAACA 76

RESULT 5
 ID ADN99172 standard; cdDNA; 1287 BP.
 XX

XX WPI; 2001-639362/73.
 DR P-PSDB; ABC07985.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID NO 7976; 103pp; English.
 XX
 DR WPI; 2004-119264/12.
 PT New human genome-derived single exon nucleic acid probes useful for human
 PT gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 XX
 PS Claim 1; SEQ ID NO 20632; 80pp; English.
 XX
 CC The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 688 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridises under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subscription, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above). The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterising gross
 CC alternative splicing events, in detecting and characterising gross
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence data for this
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030194704
 XX
 SQ Sequence 1726 BP; 602 A; 326 C; 349 G; 449 T; 0 U; 0 Other;
 SQ Sequence 1726 BP; 602 A; 326 C; 349 G; 449 T; 0 U; 0 Other;
 Query Match 78.3%; Score 18; DB 5; Length 1726;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 4 TCTTAAGGAGACCCACA 21
 Db 59 TCTTAAGGAGACCCACA 76
 XX
 RESULT 7
 ACH87337
 ID ACH87437 standard; DNA; 1766 BP.
 XX
 AC ACH87437;
 XX
 DT 29-JUL-2004 (first entry)
 XX Human genome derived single exon probe #20632.
 KW Human; Probe; SS; gene expression; single exon probe; microarray;
 KW alternative splicing event; genomic alteration.
 XX
 OS Homo sapiens.
 XX
 PN US2003194704-A1.
 XX
 PD 16-OCT-2003.
 XX
 PP 03-APR-2003; 2002US-00029386.
 XX
 PR 03-APR-2002; 2002US-00029386.
 XX (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.
 XX
 PI Penn SG, Rank DR, Hanzel DK,
 OS Homo sapiens.
 XX
 DE HTRM clone 438283 DNA sequence.
 XX
 ID AAZ52430 standard; DNA; 2077 BP.
 XX
 AC AAZ52430;
 XX
 DT 24-FEB-2000 (first entry)
 RESULT 8
 AAZ52430
 ID AAZ52430 standard; DNA; 2077 BP.
 XX
 AC AAZ52430;
 XX
 DT 24-FEB-2000 (first entry)
 XX
 DE HTRM clone 438283 DNA sequence.
 XX
 KW human transcriptional regulatory molecule; arteriosclerosis; AIDS;
 KW arteriosclerosis; cirrhosis; cancer; leukaemia; diabetes mellitus; SS;
 KW Addison's disease; multiple sclerosis; rheumatoid arthritis; infection;
 KW trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.
 XX
 OS Homo sapiens.

PR 20-NOV-1998; 98US-0109283P.
 PR 19-NOV-1999; 99US-00443041.
 XX
 PA (PAMO/) FAMODU O O.
 PA (OROZ/) OROZCO E M.
 PA (RAFA/) RAFALSKI J M.
 PA (SHEN/) SHEN J B.
 Famodu OO, Orozco EM, Rafalski JA, Shen JB;
 PI DR WPI; 2004-052051/05.
 XX P-PSDB, AD18999.
 PT New polynucleotide encoding a transcription factor polypeptide, useful in
 PT genetic mapping for plant breeding.
 XX
 PX Claim 2; SEQ ID NO 19; 43pp; English.
 XX
 CC The invention relates to a polynucleotide which encodes a transcription
 CC factor. The nucleic acid fragment is useful in genetic mapping for plant
 CC breeding. The present sequence is sterol delta-7 reductase
 CC cDNA.
 XX Sequence 1646 BP; 444 A; 343 C; 350 G; 509 T; 0 U; 0 Other;
 XX
 Query Match 77.4%; Score 17.8%; DB 12; Length 1646;
 Best Local Similarity 90.5%; Pred. No. 1.9e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 SQ
 Qy 1 GTCCTTAAGGAGCACCA 21
 Db 1460 GTCCTTAAGGAGCACCA 1440
 XX
 RESULT 11
 ADA02738/c
 ID ADA02738 standard; DNA; 52302 BP.
 XX
 AC ADA02738;
 XX DT 06-NOV-2003 (first entry)
 DE Human CCND2 carcinoma associated gene, SEQ ID NO:1256.
 XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX WO2003057146-A2.
 PN 17-JUL-2003.
 PD 26-DEC-2002; 2003WO-US041414.
 PR 26-DEC-2001; 2001US-00035832.
 PR (SAGR-) SAGRES DISCOVERY.
 PA
 PI Morris DW;
 XX
 DR WPI; 2003-587068/55.
 PT New recombinant nucleic acid encoding carcinoma associated protein,
 PT useful for preparing compositions for treating carcinomas.
 XX
 PS Claim 1; SEQ ID NO 1256; 245pp; English.
 XX
 CC The invention relates to recombinant carcinoma associated (CA) nucleic
 CC acid sequences from mouse and human (ADA0182-ADA0309), and to
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The
 CC invention also encompasses expression vectors and host cells comprising a
 CC nucleic acid, a polypeptide (especially an antibody) that specifically
 binds to the protein, and a biochip comprising CA nucleic acid, or
 fragments thereof. The sequences of the invention were identified using
 oncogenic retroviruses, which insert into the genome of the host organism
 at random. Many of these do not carry transduced host oncogenes or
 pathogenic trans-acting viral genes, meaning that cancer incidence is a
 direct consequence of the effects of proviral integration into host
 protooncogenes. The CA nucleic acid sequences can be used to diagnose
 carcinoma (especially breast cancer, prostate cancer, lymphoma or
 leukemia) or a propensity to carcinoma by determination of the sequence
 of a CA gene, or by determination of CA gene expression in particular
 tissues. CA nucleic acids, proteins and antibodies are also useful as
 therapeutic agents and in screening and evaluating drug candidates. The
 present sequence represents a specifically claimed human CA nucleic acid
 sequence of the invention. Note: The complete sequence data for this
 patent did not form part of the printed specification, but was obtained
 in electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;
 XX
 Query Match 77.4%; Score 17.8%; DB 9;
 Best Local Similarity 90.5%; Pred. No. 2.8e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 SQ
 Qy 3 GTCCTTAAGGAGCACCA 23
 Db 13323 GTCCTTAAGGAGCACCA 133303
 XX
 RESULT 12
 ADB72476/c
 ID ADB72476 standard; DNA; 52302 BP.
 XX
 AC ADB72476;
 XX DT 04-DEC-2003 (first entry)
 DE Human CCND2 gene.
 XX KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
 KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
 XX Homo sapiens.
 XX WO2003008583-A2.
 PN 30-JAN-2003.
 PD
 XX 26-DEC-2001; 2001WO-US051291.
 XX PR 02-MAR-2001; 2001US-00798586.
 PR 23-OCT-2001; 2001US-0004113.
 PR 08-NOV-2001; 2001US-00052482.
 PR 30-NOV-2001; 2001US-0099722.
 PR 20-DEC-2001; 2001US-00034650.
 XX PA (SAGR-) SAGRES DISCOVERY.
 XX PI Morris DW, Engelhard EK;
 XX DR WPI; 2003-239337/23.
 XX
 PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.
 XX
 PS Claim 1; SEQ ID NO 304; 2304pp; English.
 XX
 CC The invention relates to a novel recombinant nucleic acid comprising a
 CC nucleotide sequence selected from any of the 660 sequences fully defined
 CC in the specification. A polynucleotide of the invention has cytostatic
 CC activity, and may have a use in gene therapy, or in a vaccine. The
 CC recombinant nucleic acids and polypeptides are useful for treating
 carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
 CC

CC barcomas. The present sequence represents a human gene of the invention.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

SQ Query Match 77.4%; Score 17.8; DB 10; Length 52302;

Best Local Similarity 90.5%; Pred. No. 2.8e+02; Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OS Homo sapiens.

XX US2004072154-A1.

Db 13323 GTTTTAGGAGACCCACAGC 13303

RESULT 13
ADC85218/c

ID ADC85218 standard; DNA; 52302 BP.

XX ADC85218;

AC XX 01-JAN-2004 (first entry)

XX DE Human Conn2 genomic sequence.

XX KW Cytostatic; Gene therapy; vaccine; cancer; carcinoma-associated gene; CA;

KW secreted; transmembrane; intracellular; ds.

XX OS Homo sapiens.

XX WO200345230-A2.

XX PN 05-JUN-2003.

XX PD 02-DEC-2002; 2002WO-US038582.

XX PR 30-NOV-2001; 2001US-00997722.

XX (SAGR-) SAGRES DISCOVERY.

XX PA Morris DW, Engelhardt EK;

XX PI Morris DW, Engelhardt EK;

XX WPI; 2003-513603/48.

XX New recombinant nucleic acid comprising a nucleotide sequence of any of PT the carcinoma associated (CA) genes, useful for screening for drug candidates for diagnosing or treating carcinomas.

XX PS 01; SEQ ID NO 4; 983PP; English.

XX The invention relates to a recombinant nucleic acid comprising a nucleotide sequence selected from any of the fully defined carcinoma-associated (CA) genes from the 50 tables given in the specification. The CA proteins are secreted, transmembrane or intracellular proteins. The recombinant nucleic acids are useful for screening for drug candidates for diagnosing or treating carcinomas. Sequences given in ADC85215-ADC85514 represent CA genes of the invention.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

SQ Query Match 77.4%; Score 17.8; DB 10; Length 52302;

Best Local Similarity 90.5%; Pred. No. 2.8e+02; Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OS 3 GCTTTAGGAGACCCACAGC 23

Db 13323 GTTTTAGGAGACCCACAGC 13303

RESULT 14
ADM74333/c

ID ADM74333 standard; DNA; 52302 BP.

XX AC XX ADM74333;

XX AC XX

DT 01-JUL-2004 (first entry)

XX Human carcinoma associated (CA) nucleic acid #1.

CC Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;

CC carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma; cytostatic.

CC OS Homo sapiens.

XX XX US2004072154-A1.

Db 15-APR-2004.

XX PD 30-NOV-2001; 2001US-00997722.

XX PP 22-DEC-2000; 2000US-0074377.

XX PR 02-MAR-2001; 2001US-00798586.

XX (MORR-) MORRIS D W.

XX (ENGIE-) ENGELHARD E K.

XX PI Morris DW, Engelhardt EK;

XX DR WPI; 2004-328562/30.

XX PT New carcinoma associated gene or protein, useful for preparing a

PT composition for diagnosing or treating carcinoma e.g., leukemia or

PT lymphoma.

XX PS 1; SEQ ID NO 4; 299P; English.

XX XX

CC The invention relates to new recombinant nucleic acids. The invention

CC also relates to a host cell comprising a recombinant nucleic acid or expression vector, an expression vector comprising a recombinant nucleic acid, a recombinant protein, a method of screening for drug candidates, a

CC method of screening for a bioactive agent capable of binding to a carcinoma associated protein (CAP) encoded by nucleotide sequence, a

CC method of screening for a bioactive agent capable of modulating the activity of a CAP, a method of evaluating the effect of a candidate

CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting the activity of a CAP, a method of treating carcinoma, a method of

CC neutralising the effect of a CAP and a method of diagnosing carcinoma or propensity to carcinoma. A method of evaluating the effect of a candidate

CC carcinoma drug comprises administering the drug to a patient, removing a cell sample from the patient and determining alterations in the

CC expression or activation of a gene comprising the nucleotide sequence. A

CC method of diagnosing carcinoma comprises determining the expression of one or more genes comprising the nucleic acid sequence in a first tissue

CC type of a first individual and comparing the expression of the gene from

CC an unaffected individual, where a difference in the expression indicates

CC that the first individual has carcinoma. A method of inhibiting the activity of a CAP comprises binding an inhibitor to the CAP. Treating

CC carcinomas comprises administering to a patient an inhibitor of CAP. Neutralising the effect of a CAP comprises contacting an agent specific

CC for the CAP. The polypeptide specifically binds to the protein encoded by

CC the nucleic acid. It comprises an antibody that specifically binds to the protein encoded by the nucleic acid. The nucleic acids are useful for

CC preparing a composition for diagnosing or treating carcinoma e.g.,

CC leukaemia or lymphoma. This sequence represents a human carcinoma associated (CA) nucleic acid of the invention. Note: The sequence data for this patent did not form part of the printed specification but was

CC obtained in electronic format directly from USPTO at

XX seqdata.uspto.gov/sequence.html.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

SQ Query Match 77.4%; Score 17.8; DB 12; Length 52302;

Best Local Similarity 90.5%; Pred. No. 2.8e+02; Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GCTTTAGGAGACCCACAGC 23

Db 13323 GTTTTAGGAGACCCACAGC 13303

Qy 3 GCTTTAGGAGACCCACAGC 23

Db 133233 GTTTTAGAGACCAAGC 13303

RESULT 15
 ACN44062 Standard; DNA; 144035 BP.
 XX ID ACN44062
 XX ACN44062;
 XX DT 18-NOV-2004 (first entry)
 XX DE Human genomic sequence hCG38705.
 XX KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
 XX OS Homo sapiens.
 XX PN WO2003073826-A2.
 XX PD 12-SEP-2003.
 XX PF 28-FEB-2003; 2003WO-US006235.
 XX PR 01-MAR-2002; 2002US-00087192.
 XX PA (SAGR-) SAGRES DISCOVERY.
 XX P1 Morris DW;
 XX DR 2003-328604/31.
 XX PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 PT comprises a nucleotide sequence.
 XX PS Claim 1; SEQ ID NO 322; Opp; English.

XX The present invention relates to novel DNA and protein sequences which
 CC are associated with carcinomas. The sequences are useful for: (i) for
 CC screening drug candidates; (ii) for screening of bioactive agent capable
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the
 CC genes are useful as DNA vaccines and the CAP are useful as markers of
 CC carcinoma including lymphoma. The present sequence is one such CA coding
 CC sequence. Note: This Patent is equivalent to basic patent
 CC US20021825861, for which no sequence data was published

XX Sequence 144035 BP; 43371 A; 28888 C; 27578 G; 44198 T; 0 U; 0 Other;
 XX SQ Query Match 77.4%; Score 17.8; DB 11; Length 144035;
 XX Best Local Similarity 90.5%; Pred. No. 3.2e+02;
 XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 2 TGCTTAAAGGAGACCAAG 22
 Db 137647 TGCTTAAAGGAGACCAAG 137667

Search completed: March 9, 2005, 07:25:24
 Job time : 430 secs

This Page Blank (uspto)

Result No.	Score	Query Match	Length	DB ID	Description
1	18.8	81.7	102053	4	US-09-949-016-13025
c 2	17.8	77.4	601	4	US-09-949-016-23241
c 3	17.8	77.4	601	4	US-09-949-016-73745
c 4	17.8	77.4	1646	3	US-09-943-041A-19
c 5	17.8	77.4	36302	4	US-09-943-016-11998
c 6	17.8	77.4	36302	4	US-09-949-016-13891
c 7	17.2	74.8	1603	3	US-09-103-664A-3
c 8	17.2	74.8	1889	3	US-09-103-664A-10
c 9	17.2	74.8	2091	3	US-09-103-664A-1
c 10	17.2	74.8	2872	1	US-08-015-985-4
c 11	17.2	74.8	2872	4	US-09-285-597-4
c 12	16.8	73.0	25377	4	US-09-949-016-14261
c 13	16.8	73.0	25377	4	US-09-949-016-14261
c 14	16.8	73.0	28791	4	US-09-949-016-15396
c 15	16.6	72.2	1605	4	US-09-023-655-194
c 16	16.6	72.2	49407	4	US-09-949-016-12532
c 17	16.6	72.2	49408	4	US-09-949-016-17045
c 18	16.6	72.2	301828	4	US-09-949-016-13969
c 19	16.4	71.3	601	4	US-09-949-016-29750
c 20	16.4	71.3	601	4	US-09-949-016-63179
c 21	16.4	71.3	38239	4	US-09-949-016-12348
c 22	16.4	71.3	38232	4	US-09-949-016-13570
c 23	16.4	71.3	670689	4	US-09-949-016-12505
c 24	16.4	71.3	670650	4	US-09-949-016-14207
c 25	16.2	70.4	25	4	US-09-396-196G-42041
c 26	16.2	70.4	25	4	US-09-396-196G-42042
c 27	16.2	70.4	25	4	US-09-396-196G-42044

```

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fasleseq For Windows Version 4.0
; SEQ ID NO: 73741
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-23241

Query Match
Best Local Similarity 90.5%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Length 601;
RESULT 5
US-09-949-016-11998/C
Sequence 11998, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY NUMBER: 60/241,755
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 11998
; LENGTH: 36302
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) .(36302)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11998

Query Match
Best Local Similarity 90.5%; Pred. No. 64;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Length 36302;
RESULT 5
US-09-949-016-13891/C
Sequence 13891, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY NUMBER: 60/241,755
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 13891

Query Match
Best Local Similarity 90.5%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Length 36302;
RESULT 5
US-09-949-016-13891/C
Sequence 13891, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY NUMBER: 60/241,755
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 13891

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Microsoft Office 97
US-09-943-041A-19/C
Sequence 19, Application US/09443041A
; Patent No. 646517
; GENERAL INFORMATION:
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Ratajski, Antoni
; APPLICANT: Shen, Jennie
; TITLE OF INVENTION: Sterol Metabolism Enzymes
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/443, 041A
; CURRENT FILING DATE: 1999-11-18
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
US-09-943-041A-19
Sequence 19, Application US/09443041A
; Patent No. 646517
; GENERAL INFORMATION:
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Ratajski, Antoni
; APPLICANT: Shen, Jennie
; TITLE OF INVENTION: Sterol Metabolism Enzymes
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/443, 041A
; CURRENT FILING DATE: 1999-11-18
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97

```

LENGTH: 36302
 TYPE: DNA
 ORGANISM: Human
 FEATURE: misc_feature
 NAME/KEY: misc_feature
 LOCATION: (1)..(36302)
 OTHER INFORMATION: n = A,T,C or G
 US-09-949-016-13891

Qy 2 TGCTTAAAGGAGACCAACAGC 23
 Db 1732 TGGTTAGGAGACCAACAGC 1711

RESULT 9
 US-09-103-664A-1/c
 ; Sequence 1, Application US/09103664A
 ; Patent No. 6453358
 ; GENERAL INFORMATION:
 ; APPLICANT: University of Kentucky Research Foundation
 ; APPLICANT: Timoney, John
 ; APPLICANT: Artiushin, Sergey
 ; TITLE OF INVENTION: Compounds Encoding the Protective M-Like Protein of Streptococcus
 ; TITLE OF INVENTION: and Assays Therefor
 ; FILE REFERENCE: 50229-212
 ; CURRENT APPLICATION NUMBER: US/09/103, 664A
 ; CURRENT FILING DATE: 1998-06-23
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 2091
 ; TYPE: DNA
 ; ORGANISM: Streptococcus equi
 US-09-103-664A-1

Query Match Score 17.8; DB 4; Length 36302;
 Best Local Similarity 90.5%; Pred. No. 64; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GCTCTAAAGGAGACCAACAGC 23
 Db 5323 GTTTAAAGGAGACCAACAGC 5303

RESULT 7
 US-09-103-664A-3/c
 ; Sequence 3, Application US/09103664A
 ; Patent No. 6453358
 ; GENERAL INFORMATION:
 ; APPLICANT: University of Kentucky Research Foundation
 ; APPLICANT: Timoney, John
 ; APPLICANT: Artiushin, Sergey
 ; TITLE OF INVENTION: Compounds Encoding the Protective M-Like Protein of Streptococcus
 ; TITLE OF INVENTION: and Assays Therefor
 ; FILE REFERENCE: 50229-21.2
 ; CURRENT APPLICATION NUMBER: US/09/103, 664A
 ; CURRENT FILING DATE: 1998-06-23
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 1.603
 ; TYPE: DNA
 ; ORGANISM: Streptococcus equi
 US-09-103-664A-3

Query Match Score 17.8; DB 4; Length 36302;
 Best Local Similarity 90.5%; Pred. No. 64; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TGCTTAAAGGAGACCAACAGC 23
 Db 1594 TGGTTAGGAGACCAACAGC 1573

RESULT 10
 US-08-015-985-4
 ; Sequence 4, Application US/08015985
 ; Patent No. 5538886
 ; GENERAL INFORMATION:
 ; APPLICANT: Schlessinger, Joseph
 ; APPLICANT: Sap, Jan M.
 ; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PENNIE & EDMONDS
 ; STREET: 1155 AVENUE OF THE AMERICAS
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER: READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/015, 985
 ; FILING DATE: 10-FEB-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Coruzzi, Laura A.
 ; REGISTRATION NUMBER: 30,742
 ; REFERENCE/DOCKET NUMBER: 7683-020
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9690
 ; TELEFAX: (212) 869-9741/8864
 ; TELEX: 66141 PENNIE
 ; SEQUENCE FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2872 base pairs
 ; TYPE: nucleic acid

Query Match Score 17.2; DB 3; Length 1889;
 Best Local Similarity 86.4%; Pred. No. 87; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGGTTAGGAGACCAACAGC 23
 Db 1594 TGGTTAGGAGACCAACAGC 1573

RESULT 8
 US-09-103-664A-10/c
 ; Sequence 10, Application US/09103664A
 ; Patent No. 6453358
 ; GENERAL INFORMATION:
 ; APPLICANT: University of Kentucky Research Foundation
 ; APPLICANT: Timoney, John
 ; APPLICANT: Artiushin, Sergey
 ; TITLE OF INVENTION: Compounds Encoding the Protective M-Like Protein of Streptococcus
 ; TITLE OF INVENTION: and Assays Therefor
 ; FILE REFERENCE: 50229-21.2
 ; CURRENT APPLICATION NUMBER: US/09/103, 664A
 ; CURRENT FILING DATE: 1998-06-23
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 10
 ; LENGTH: 1889
 ; TYPE: DNA
 ; ORGANISM: Streptococcus equi
 US-09-103-664A-10

Query Match Score 17.2; DB 3; Length 1889;
 Best Local Similarity 86.4%; Pred. No. 87; Mismatches 3; Indels 0; Gaps 0;

```

; GENERAL INFORMATION: ; APPLICANT: VENTER, J. Craig et al. ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; TOPOLYGE: unknown ; FILE REFERENCE: CL001307
; MOLECULE TYPE: cDNA ; CURRENT APPLICATION NUMBER: US/09/949,016
; US-08-015-985-4 ; CURRENT FILING DATE: 2000-04-14
; ; PRIORITY APPLICATION NUMBER: 60/241,755
; ; PRIORITY FILING DATE: 2000-10-20
; ; PRIORITY APPLICATION NUMBER: 60/237,768
; ; PRIORITY FILING DATE: 2000-10-03
; ; PRIORITY APPLICATION NUMBER: 60/231,498
; ; PRIORITY FILING DATE: 2000-09-08
; ; NUMBER OF SEQ ID NOS: 207012
; ; SOFTWARE: FastSEQ for Windows Version 4.0
; ; SEQ ID NO: 14260
; ; LENGTH: 25377
; ; TYPE: DNA
; ; ORGANISM: Human
; ; US-09-949-016-14260

Query Match 74.8%; Score 17.2; DB 1; Length 2872;
Best Local Similarity 86.4%; Pred. No. 92;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 GGTCTTAAAGGAGACCAAG 22
Db 2188 GGTCTTAAAGGAGACATCACAG 2209

RESULT 11
US-09-280-597-4
Sequence 4, Application US/09280597
Patent No. 6632905
GENERAL INFORMATION:
APPLICANT: Schlesinger, Joseph
ADDRESS: SAP, Jan M.
TITLE OF INVENTION: NOVEL RECEPTOR TYPE PHOSPHOTYROSINE
TITLE OF INVENTION: PHOSPHATASE-ALPHA
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
STREET: PENNIE & EDMONDS
CITY: NEW YORK
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/280,597
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/015,985
FILING DATE: 10-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7603-020
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-09-280-597-4

Query Match 74.8%; Score 17.2; DB 4; Length 2872;
Best Local Similarity 86.4%; Pred. No. 92;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 GGTCTTAAAGGAGACCAAG 22
Db 2188 GGTCTTAAAGGAGACATCACAG 2209

RESULT 12
US-09-949-016-14260
Sequence 14260
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307

Query Match 73.0%; Score 16.8; DB 4; Length 25377;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GGTCTTAAAGGAGACCAAC 20
Db 12407 GGTCTTCAAGGAGACACTAC 12426

RESULT 13
US-09-949-016-14261
Sequence 14261, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIORITY APPLICATION NUMBER: 60/241,755
PRIORITY FILING DATE: 2000-10-20
PRIORITY APPLICATION NUMBER: 60/237,768
PRIORITY FILING DATE: 2000-10-03
PRIORITY APPLICATION NUMBER: 60/231,498
PRIORITY FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO: 14261
LENGTH: 25377
TYPE: DNA
ORGANISM: Human
US-09-949-016-14261

Query Match 73.0%; Score 16.8; DB 4; Length 25377;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GGTCTTAAAGGAGACCAAC 20
Db 12407 GGTCTTCAAGGAGACACTAC 12426

RESULT 14
US-09-949-016-15396/c
Sequence 15396, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307

```

CURRENT APPLICATION NUMBER: US/09/949,016
 CURRENT FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 20/012
 SEQ ID NO: 15396
 SOFTWARE: FastSEQ for Windows Version 4.0
 LENGTH: 28791
 TYPE: DNA
 ORGANISM: Human
 US-09-949-016-15396

Query Match 73.0%; Score 16.8; DB 4; Length 28791;
 Best Local Similarity 90.0%; Pred. No. 1.9e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TGTCTTAAGGAGACCAACA 21
 Db 3097 TGCTTAAAGAGACGACACA 3078

RESULT 15
 US-09-023-655-194/c
 Sequence 194, Application US/09023655
 Patent No. 6603879

GENERAL INFORMATION:
 APPLICANT: Cocks, Benjamin G.
 APPLICANT: Susan G. Stuart
 APPLICANT: Jeffrey J. Seilhamer
 TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
 NUMBER OF SEQUENCES: 1508
 NUMBER OF INVENTION: EXPRESSION
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 STREET: 3174 PORTER DRIVE
 CITY: PALO ALTO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 94304

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/023,655
 FILING DATE: HEREWITH

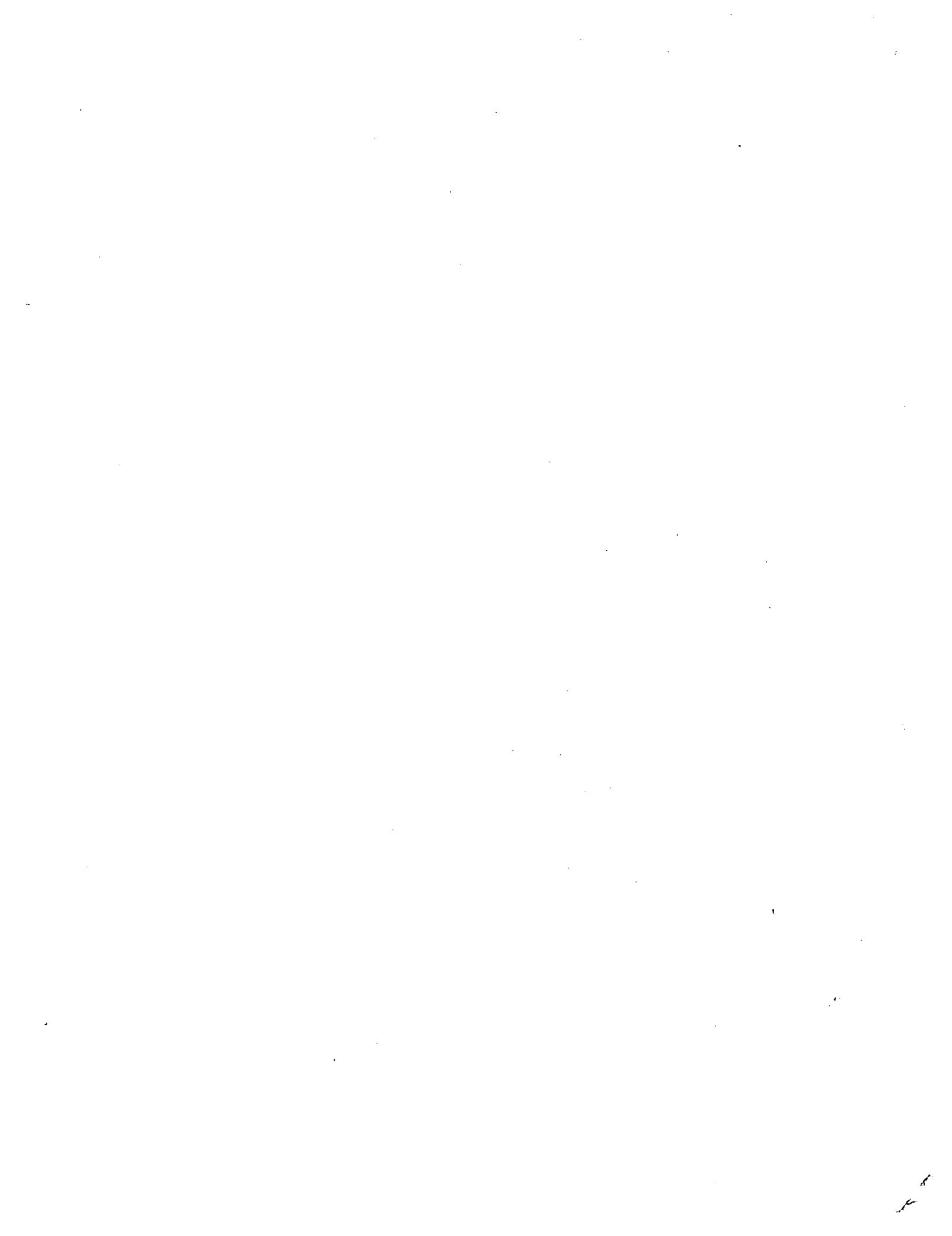
CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Zeller, Karen J.
 REGISTRATION NUMBER: 37,071
 REFERENCE/DOCKET NUMBER: PA-0001 US

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 855-0555
 TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 194:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1605 base pairs
 TYPE: nucleic acid
 STRANDBNESS: single
 TOPOLOGY: linear

IMMEDIATE SOURCE:
 LIBRARY: THP1PEB01
 CLONE: 073293

US-09-023-655-194



Sequence 20632, Application US/10029386
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharron G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 FILE REFERENCE: AEONICA-X2
 CURRENT APPLICATION NUMBER: US/10/029,386
 CURRENT FILING DATE: 2001-12-20
 NUMBER OF SEQ ID NOS: 34288
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 20632
 LENGTH: 1766
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC010328.4
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3
 OTHER INFORMATION: EST HUMAN HIT: BE253605.1, EVALUATE 0.00e+00
 OTHER INFORMATION: SWISSPROT HIT: PS1522, EVALUATE 0.00e+00
 OTHER INFORMATION: NT HIT: 918922743, EVALUATE 1.00e-128
 US-10-029-386-20632

Query Match, 78.3%; Score 18; DB 16; Length 1766;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 TCTTAAGGAGACACCA 21
 Db 538 TCTTAAGGAGACACCA 555

RESULT 3
 US-10-224-880C-19/c
 Sequence 19, Application US/10224880C
 GENERAL INFORMATION:
 APPLICANT: Famodu, Omolayo O.
 APPLICANT: Orozco, Emil M.
 APPLICANT: Rafalski, J. Antoni
 APPLICANT: Shen, Jennie B.
 TITLE OF INVENTION: Sterol Metabolism Enzymes
 FILE REFERENCE: BBJ1282 US DIV
 CURRENT APPLICATION NUMBER: US/10/224,880C
 CURRENT FILING DATE: 2002-08-21
 PRIOR APPLICATION NUMBER: 60/109,283
 PRIOR FILING DATE: 1998-11-20
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 19
 LENGTH: 1646
 TYPE: DNA
 ORGANISM: Glycine max
 US-10-224-880C-19

Query Match, 77.4%; Score 17.8; DB 17; Length 1646;
 Best Local Similarity 90.5%; Pred. No. 67;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTGTCCTAAGGAGACACCA 21
 Db 1460 CTGTCCTAAGGAGACACCA 1440

RESULT 4
 US-10-087-192-322

Query Match, 77.4%; Score 17.8; DB 11; Length 52302;
 Best Local Similarity 90.5%; Pred. No. 87;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GTCTTAAAGGAGACACAGC 23
 Db 13323 GTTTTAAAGGAGACACAGC 13303

RESULT 5
 US-09-997-722-4/c
 Sequence 4, Application US/09997722
 GENERAL INFORMATION:
 APPLICANT: Morris, David
 APPLICANT: Engelhard, Eric
 TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
 FILE REFERENCE: A-71171/RMS/DCP
 CURRENT APPLICATION NUMBER: US/09/997,722
 CURRENT FILING DATE: 2001-11-30
 PRIOR APPLICATION NUMBER: US 09/747,377
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: US 09/798,386
 PRIOR FILING DATE: 2001-03-02
 NUMBER OF SEQ ID NOS: 301
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 4
 LENGTH: 52302
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (11301)..(11320)
 OTHER INFORMATION: "n" at positions 11301 through 11320 can be any base.
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (29267)..(29286)
 OTHER INFORMATION: "n" at positions 29267 through 29286 can be any base.
 US-09-997-722-4

Query Match, 77.4%; Score 17.8; DB 11; Length 52302;
 Best Local Similarity 90.5%; Pred. No. 87;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GTCTTAAAGGAGACACAGC 23
 Db 13323 GTTTTAAAGGAGACACAGC 13303

RESULT 6
 US-10-087-192-322

```

Sequence 322, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
CANCER
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-01-01
PRIOR APPLICATION NUMBER: US/9747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SEQ ID NO 322
TYPE: DNA
LENGTH: 144035
ORGANISM: Homo sapiens
us-10-087-192-322

Query Match 77.4%; Score 17.8; DB 13; Length 144035;
Best Local Similarity 90.5%; Prd. No. 93; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2;
Qy 2 TGTCTTAAAGAGACCAAG 22
Db 137647 TGCTTGAAGAGACCAAG 137667

RESULT 7
US-10-741-601-9071/C
Sequence 9071, Application US/10741601
Publication No. US20040166519A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
STENOSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001500
CURRENT APPLICATION NUMBER: US/10/741,601
CURRENT FILING DATE: 2003-12-22
NUMBER OF SEQ ID NOS: 26415
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 9071
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
us-10-741-601-9071

Query Match 75.7%; Score 17.4; DB 18; Length 201;
Best Local Similarity 94.7%; Prd. No. 92; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1;
Qy 4 TCTTAAGAGACCAAG 22
Db 22 TCTTCAAGAGACCAAG 4

RESULT 8
US-10-741-601-9089/C
Sequence 9089, Application US/10741601
Publication No. US20040166519A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
STENOSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001500
CURRENT APPLICATION NUMBER: US/10/741,601
CURRENT FILING DATE: 2003-12-22
NUMBER OF SEQ ID NOS: 26415
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 9089

Query Match 75.7%; Score 17.4; DB 18; Length 201;
Best Local Similarity 94.7%; Prd. No. 92; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1;
Qy 4 TCTTAAGAGACCAAG 22
Db 180 TCTTCAAGAGACCAAG 198

RESULT 9
US-10-741-601-9487/C
Sequence 9487, Application US/10741601
Publication No. US20040166519A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
STENOSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001500
CURRENT APPLICATION NUMBER: US/10/741,601
CURRENT FILING DATE: 2003-12-22
NUMBER OF SEQ ID NOS: 26415
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 9487
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
us-10-741-601-9487

Query Match 75.7%; Score 17.4; DB 18; Length 201;
Best Local Similarity 94.7%; Prd. No. 92; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1;
Qy 4 TCTTAAGAGACCAAG 22
Db 40 TCTTCAAGAGACCAAG 22

RESULT 10
US-10-741-601-12984
Sequence 12984, Application US/10741601
Publication No. US20040166519A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
STENOSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001500
CURRENT APPLICATION NUMBER: US/10/741,601
CURRENT FILING DATE: 2003-12-22
NUMBER OF SEQ ID NOS: 26415
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 12984
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
us-10-741-601-12984

Query Match 75.7%; Score 17.4; DB 18; Length 201;
Best Local Similarity 94.7%; Prd. No. 92; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1;
Qy 4 TCTTAAGAGACCAAG 22
Db 180 TCTTCAAGAGACCAAG 198

RESULT 11
US-10-741-601-12992
Sequence 12992, Application US/10741601

```

Publication No. US20040166519A1
 GENERAL INFORMATION: CARGILL, Michelle et al.
 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 FILE REFERENCE: CL001500
 CURRENT APPLICATION NUMBER: US/10/741,601
 CURRENT FILING DATE: 2003-12-22
 NUMBER OF SEQ ID NOS: 26415
 SOFTWARE: Fast-SEQ For Windows Version 4.0
 LENGTH: 201
 TYPE: DNA
 ORGANISM: Homo sapiens

US-10-741-601-12932

Query Match 75.7%; Score 17.4; DB 18; Length 201;
 Best Local Similarity 94.7%; Pred. No. 92;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TCTTAAGGAGACCAACAG 22
 Db 130 TCTTCAGGAGACCAACAG 148

RESULT 12
 US-10-741-601-13191
 Sequence 13191, Application US/10741601
 Publication No. US20040166519A1
 GENERAL INFORMATION:
 APPLICANT: CARGILL, Michelle et al.
 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 FILE REFERENCE: CL001500
 CURRENT APPLICATION NUMBER: US/10/741,601
 CURRENT FILING DATE: 2003-12-22
 NUMBER OF SEQ ID NOS: 26415
 SOFTWARE: Fast-SEQ For Windows Version 4.0
 LENGTH: 201
 TYPE: DNA
 ORGANISM: Homo sapiens

US-10-741-601-13191

Query Match 75.7%; Score 17.4; DB 18; Length 201;
 Best Local Similarity 94.7%; Pred. No. 92;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TCTTAAGGAGACCAACAG 22
 Db 162 TCTTCAGGAGACCAACAG 180

RESULT 13
 US-10-741-600-23447/c
 Sequence 23447, Application US/10741600
 Publication No. US20050026165A1
 GENERAL INFORMATION:
 APPLICANT: CARGILL, Michelle et al.
 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 FILE REFERENCE: CL001499
 CURRENT APPLICATION NUMBER: US/10/741,600
 CURRENT FILING DATE: 2003-12-22
 NUMBER OF SEQ ID NOS: 73997
 SOFTWARE: Fast-SEQ For Windows Version 4.0
 SEQ ID NO 23447
 LENGTH: 201
 TYPE: DNA
 ORGANISM: Homo sapiens

US-10-741-600-23447

Query Match 75.7%; Score 17.4; DB 19; Length 201;
 Best Local Similarity 94.7%; Pred. No. 92;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TCTTAAGGAGACCAACAG 22
 Db 40 TCTTCAGGAGACCAACAG 22

RESULT 14
 US-10-741-600-23447/c
 Sequence 23447, Application US/10741600
 Publication No. US20050026165A1
 GENERAL INFORMATION:
 APPLICANT: CARGILL, Michelle et al.
 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 FILE REFERENCE: CL001499
 CURRENT APPLICATION NUMBER: US/10/741,600
 CURRENT FILING DATE: 2003-12-22
 NUMBER OF SEQ ID NOS: 73997
 SOFTWARE: Fast-SEQ For Windows Version 4.0
 SEQ ID NO 23447
 LENGTH: 201
 TYPE: DNA
 ORGANISM: Homo sapiens

US-10-741-600-23447

Query Match 75.7%; Score 17.4; DB 19; Length 201;
 Best Local Similarity 94.7%; Pred. No. 92;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TCTTAAGGAGACCAACAG 22
 Db 72 TCTTCAGGAGACCAACAG 54

RESULT 15
 US-10-741-600-23845/c
 Sequence 23845, Application US/10741600
 Publication No. US20050026165A1
 GENERAL INFORMATION:
 APPLICANT: CARGILL, Michelle et al.
 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 FILE REFERENCE: CL001499
 CURRENT APPLICATION NUMBER: US/10/741,600
 CURRENT FILING DATE: 2003-12-22
 NUMBER OF SEQ ID NOS: 73997
 SOFTWARE: Fast-SEQ For Windows Version 4.0
 SEQ ID NO 23845
 LENGTH: 201
 TYPE: DNA
 ORGANISM: Homo sapiens

US-10-741-600-23845

Query Match 75.7%; Score 17.4; DB 19; Length 201;
 Best Local Similarity 94.7%; Pred. No. 92;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TCTTAAGGAGACCAACAG 22
 Db 162 TCTTCAGGAGACCAACAG 180

Search completed: March 9, 2005, 09:01:41
 Job time : 510 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2005, 05:25:31 ; Search time 3237 Seconds (without alignment)

270,460 Million cell updates/sec

Title: US-09-835-976B-115

Perfect score: 23

Sequence: gtgtcttaaggagacaccacagg 23

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing First 45 summaries

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_htc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gsb1:*

9: gb_gsb2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	19.8	86.1	911	4 BI763138	BI763138 603050060
c 2	18.9	82.6	745	9 AG122978	AG422978 Mus muscu
c 3	18.8	81.7	811	8 BH084168	BH084168 RPCI-24-3
c 4	18.8	81.7	959	4 BG033867	BG033867 602301151
c 5	18.8	81.7	1117	9 CL01784	CL01784 CH21-486
c 6	18.8	81.7	1182	8 CC266894	CC266894 CH261-154
c 7	18.4	80.0	214	4 BG100434	BG100434 luxB5906
c 8	18.2	79.1	338	2 BE713998	BE713998 MR3-H701
c 9	18.2	79.1	631	5 BU302275	BU302275 603741788
c 10	18.2	79.1	640	1 AI981983	AI981983 pat. pk007
c 11	18.2	79.1	657	8 AQ536886	AQ536886 RPCI-11.3
c 12	18.2	79.1	698	5 BQ210941	BQ210941 UT-R-DT1-
c 13	18.2	79.1	786	7 CK9454697	CK9454697 91519 MA
c 14	18.2	79.1	905	4 BG743436	BG743436 602634952
c 15	18.2	79.1	1264	2 BB92595	BB92595 60165938
c 16	18	78.3	360	7 CF135044	CF135044 UT-HF-CB0
c 17	17.8	77.4	209	2 BB81011	BB81011 GM701013A
c 18	17.8	77.4	228	7 CF920727	CF920727 GM701013A
c 19	17.8	77.4	275	2 BB821695	BB821695 GM7001015A
c 20	17.8	77.4	323	2 AW799992	AW799992 PM4-UM005
c 21	17.8	77.4	412	4 BB81935	BB81935 sa91f08.
c 22	17.8	77.4	443	2 BB684331	BB684331 B0610855
c 23	17.8	77.4	455	5 BQ610855	B0610855 sa91f08.
	24	17.8	77.4	497	1 AU083746 ORIGIN

ALIGNMENTS

25	17.8	77.4	501	9 CIE260387	CIE260387 tiger-988-
c 26	17.8	77.4	620	5 BX257728	BX257728 CR832914
c 27	17.8	77.4	676	9 CR832914	CR832914 GROAA66B
c 28	17.8	77.4	737	8 AQ425552	AQ425552 CTIBI-EI-
c 29	17.8	77.4	756	9 CR8407120	CR8407120 tiger-988-
c 30	17.8	77.4	759	7 CR8604020	CR8604020 AGENCOURT
c 31	17.8	77.4	775	7 CRN046985	CRN046985 V2_P10_F1
c 32	17.8	77.4	847	7 CN047495	CN047495 V2_D13_G1
c 33	17.8	77.4	969	4 BG392149	BG392149 60410252
c 34	17.4	75.7	460	8 A2013466	A2013466 RPCI-23-3
c 35	17.4	75.7	534	9 CIE444879	CIE444879 tiger-988-
c 36	17.4	75.7	535	4 B1969912	B1969912 GM3009B
c 37	17.4	75.7	546	8 A2986415	A2986415 2M0268C08
c 38	17.4	75.7	550	5 BQ273623	BQ273623 sa914e06
c 39	17.4	75.7	555	2 AW96631	AW96631 sa980e09.y
c 40	17.4	75.7	573	2 AW760528	AW760528 CR194471
c 41	17.4	75.7	664	9 CR194471	CR194471 Reverse_S
c 42	17.4	75.7	746	5 BQ863139	BQ863139 QGC23B05.
c 43	17.4	75.7	756	9 AG500148	AG500148 Mus muscu
c 44	17.4	75.7	760	9 AG579385	AG579385 Mus muscu
c 45	17.4	75.7	761	9 AG460898	AG460898 Mus muscu

BI763138/c	LOCUS	BI763138	911 bp	mRNA	linear EST 25-SEP-2001
	DEFINITION	6030500601 NIH_MGC_116	Homo sapiens	cdDNA clone IMAGE:5190183_5'	
	mRNA sequence.				
	BI763138				
	ACCESSION	BI763138.1			
	VERSION	EST.			
	KEYWORDS	Homo sapiens (human)			
	SOURCE	Homo sapiens			
	ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo; Homo.			
	REFERENCE	1 (bases 1 to 911)			
	AUTHORS	NIH-MGC			
	TITLE	http://mgc.ncbi.nlm.nih.gov/			
	JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)			
	COMMENT	Unpublished (1999)			
	CONTACT	Contact: Robert Straussberg, Ph.D.			
	EMAIL	csgapbs.r@mail.nih.gov			
	Tissue	Procurement: Life Technologies, Inc.			
	CDNA	CDNA Library Preparation: Life Technologies, Inc.			
	Library	Arrayed by: The I.M.A.G.E. Consortium (L1LN1)			
	DNA	Sequencing by: Incyte Genomics, Inc.			
	Sequencing	found through the I.M.A.G.E. Consortium information can be			
		http://image.lnl.gov			
	Plate	LIAM11475	row: 9	column: 16	
		High quality sequence stop: 641.			
	Location/Qualifiers	1. 1			
		/organism="Homo sapiens"			
		/mol_type="mRNA"			
		/db_xref="taxon:9607"			
		/clone="IMAGE:5190183"			
		/lab_host="DH10B"			
		/clone_id="NIH_MGC_116"			
		/note="Organ: Pooled colon, kidney, stomach; Vector: RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomach, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."			

Query Match 86.1%; Score 19.8; DB 4; Length 911;
 Best Local Similarity 91.3%; Pred. No. 1.2e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 ORGANISM Mus musculus (house mouse)

ACCESSION BH084168
 VERSION BH084168.1
 KEYWORDS GI:14903765
 SOURCE GSS
 ORGANISM Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 811)
 AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
 Teegay,G., Geer,K., Kroli,M., Shvartsbeyn,A., Gebregeorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-24
 COMMENT Unpublished (1999)
 Other GSS: RPCI-24-366C1.TJ
 Contact: Shuying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 366 row: C column: 1
 Seq primer: T7
 Class: BAC ends.

ACCESSION AG422978
 DEFINITION Mus musculus molossinus DNA, clone:MSMg01-294P02.T7, genomic survey
 sequence.

ACCESSION AG422978
 DEFINITION AG422978.1 GI:48066041
 KEYWORDS GSS.
 SOURCE Mus musculus molossinus
 ORGANISM Mus musculus molossinus
 Mus musculus molossinus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Butheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
 TITLE BAC end Sequences of Library MSMg01
 JOURNAL Unpublished
 2 (bases 1 to 745)
 AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-722 Sushiro-chou, Tsurumi-ku, Yokohama, Kanagawa 220-0045, Japan
 (E-mail: hattori@gsc.riken.jp, URL:<http://hsp.gsc.riken.go.jp/>,
 Tel: 81-45-53-9111, Fax: 81-45-503-9170)
 COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC
 library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
 Tsukuba Institute, Bio Resource Center,
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1
 Koyadai, Tsukuba, 305-0074, Japan
 Phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@rtc.riken.jp
 PRIMERS Sequencing : T7
 LIBRARY Vector : pBACe3 .6
 R.Site 1 : ECORI
 R.Site 2 : ECORI
 FEATURES source Location/Qualifiers 1.745
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /sub_species="molossinus"
 /db_xref="taxon:57488"
 /clone="MSMg01-294P02.T7"
 /sex="male"
 /tissue_type="mixture of kidney and spleen"
 /clone_Lib=MSMg01 Mouse Male BAC Library"

ACCESSION BG033867/C
 DEFINITION mRNA sequence.

Query Match 82.6%; Score 19; DB 9; Length 745;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 ORGANISM Homo sapiens (human)

ACCESSION BG033867
 VERSION BG033867.1
 KEYWORDS GI:12426586
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 959)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov

ACCESSION BH084168/C
 DEFINITION BH084168
 LOCUS RPCI-24-366C1.TVB
 DEFINITION RPCI-24-366C1.TVB RPCI-24 Mus musculus genomic clone RPCI-24-366C1,
 Genomic survey sequence.

ACCESSION BH084168
 DEFINITION BH084168
 LOCUS GSS 18-JUL-2001
 DEFINITION GSS 18-JUL-2001 linear

Tissue Procurement: DCTD/DTP
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM10111 row: f column: 14
 High quality sequence stop: 756.

FEATURES source
 1..959
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="Taxon:9606"
 /clones="IMAGE:4402573"
 /tissue_type="mammary adenocarcinoma, cell line"
 /lab_host="PH10B (phage-resistant)"
 /clone_id="NIH MGC:87"
 /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo dT-primed.
 Average insert size 1.383 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

ORIGIN

Query Match 81.7%; Score 18.8; DB 4; Length 959;
 Best Local Similarity 90.9%; Pred. No. 3.7e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 632 GAGTCATAAGGAGACCCACAG 611

RESULT 5

CL018784/C CH216-4B6_Sp6.1 CH216_Xenopus tropicalis genomic clone CH216-4B6,
 genomic survey sequence.

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES source

1..1117

/organism="Xenopus tropicalis"

/mol_type="genomic DNA"

/strain="Nigerian frog"

/db_xref="Taxon:8364"

/clone_id="CH216-4B6"

/sex="male"

/cell_line="Stock 248 F7A2, inbred N7"

/clone_id="CH216"

/note="Vector: PTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"

ORIGIN

Query Match 81.7%; Score 18.8; DB 9; Length 1117;

Best Local Similarity 90.9%; Pred. No. 3.8e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAGTCATAAGGAGACCCACAG 22

Db 539 GAGTCATAAGGAGACCCACAG 518

RESULT 6

CC266894 LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES source

1..1182

/organism="Gallus gallus"

/mol_type="genomic DNA"

/strain="Red Jungle Fowl"

/db_xref="Taxon:9031"

/clone_id="CH216-154M10"

/sex="female"

/cell_line="UCD001", inbred

2556

/clone_id="CH216"

/note="Vector: PTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI; CH216. Female Chicken library - For library and clone ordering information: http://www.chori.org/bcpcac"

ORIGIN

Query Match 81.7%; Score 18.8; DB 8; Length 1182;

Best Local Similarity 90.9%; Pred. No. 3.8e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAGTCATAAGGAGACCCACAG 22

Db 152 GAGTCATAAGGAGACCCACAG 173

RESULT 7

BG100434/C LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES source

1..214

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/strain="Eddy type B spermatogonia"

/db_xref="Taxon:8364"

/clone_id="CH216-1B6"

/cell_line="Stock 248 F7A2, inbred N7"

/note="Vector: PTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"

ORIGIN

Query Match 81.7%; Score 18.8; DB 9; Length 214

Best Local Similarity 90.9%; Pred. No. 3.8e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAGTCATAAGGAGACCCACAG 22

Db 152 GAGTCATAAGGAGACCCACAG 173

REFERENCE AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus . (bases 1 to 214)

Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Waterston, R., and Wilson, R., McCann, R., TITLE Sequence tags of the human transcriptome with ORF expressed

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL PubMed 10202663

COMMENT Contact: Simpson A.J.G.

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR3-HR0713-0806-023-503&t3=2000-06-08&t4=1>)

FEATURES source /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_id="HR0713"

/note="Organ: head, neck; Vector: puc18; Site: 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent Application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the PUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES source /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_id="HR0713"

/note="Organ: head, neck; Vector: puc18; Site: 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent Application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the PUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES source /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_id="HR0713"

/note="Organ: head, neck; Vector: puc18; Site: 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent Application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the PUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN Query Match 79.1%; Score 18.2%; DB 2; Length 338;

Best Local Similarity 87.0%; Pred. No. 6.6e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGTCTTAAGGAGACACAGC 23

Db 70 GTGTCTACGGAGACAGACAGC 92

RESULT 9

BU302275/c

LOCUS BU302275 631 bp mRNA linear EST 27-NOV-2002

DEFINITION 603741788f1 CSEQCTANS6 Gallus gallus cDNA clone CHEST640022 5', mRNA sequence.

DEFINITION 1 (bases 1 to 631)

ACCESSION BU302275

VERSION BU302275.1

VERSION GI:25751924

EST.

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Butelostomi; Archosaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.

AUTHORS

TITLE A Comprehensive Collection of Chicken cDNAs

JOURNAL Curr. Biol. 12 (22), 1365-1369 (2002)

MEDLINE 22335534

PUBLISHED 12443592

COMMENT Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology (UMIST)

REFERENCE AUTHORS 1 (bases 1 to 338)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.,

PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008330
 Fax: 01612360409
 Email: Simon.Rubbard@umist.ac.uk.

FEATURES
 source
 Location/Qualifiers
 1. .631

/organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton Line 151"
 /db_xref="taxon:9031"
 /clone="CHEST640o22"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_id="CSB0CHN56"

/note="Organ: small intestine; Vector: pBluescript II KS (+); Site 1: EcoRI; Site 2: Not1; This is a normalized library was constructed from 1 million independent cDNA clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adaptors, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer annealing hybridization was used."

ORIGIN

Query Match Score 18.2; DB 5; Length 631;
 Best Local Similarity 87.0%; Pred. No. 7e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTGTCTTAAGGGACACCAAGC 23
 Db 265 GTGTCTCTAGGGACACAAAGC 243

RESULT 10
 AI981983

LOCUS
 DEFINITION
 pat.pk0070.f9 f. chicken activated T cell cDNA Gallus gallus cDNA
 Clone pat.pk0070.f9.f 5', mRNA sequence.

ACCESSION
 AI981983
 VERSION
 AI981983.1
 EST
 SOURCE
 Gallus gallus (chicken)

ORGANISM
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 Archosauria; Aves; Neognathae; Galiiformes; Phasianidae;
 Phasianinae; Gallus.

1. (bases 1 to 640)
 Tirunagaru, V.G., Sofer, L., Cui, J. and Burnside, J.
 An expressed sequence tag database of T-cell-enriched activated chicken splenocytes: sequence analysis of 5251 clones
 Genomics 66 (2), 144-151 (2000)

JOURNAL
 20318616
 PUBMED
 10860659
 COMMENT
 Contact: Joan Burnside
 Molecular Endocrinology
 University of Delaware
 40 Townsend Hall, Newark, DE 19717, USA
 Tel: 302 831-1345
 Fax: 302 831-3411
 Email: joan@udel.edu, www.chickest.udel.edu
 Seq primer: T7.

FEATURES
 source
 Location/Qualifiers
 1. .640

/organism="Gallus gallus"
 /mol_type="mRNA"
 /db_xref="taxon:9031"

/clone="pat.pk0070.f9."
 /sex="male"
 /cell_type="Con A-activated splenic T cell"
 /lab_host="E.coli 'TOP10 F'"
 /clone_id="chicken activated T cell cDNA"
 /note="Vector: pcDNA3"

ORIGIN

Query Match Score 18.2; DB 1; Length 640;
 Best Local Similarity 87.0%; Pred. No. 7.1e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTGTCTTAAGGGACACCAAGC 23
 Db 403 GTGTCTTAAGGGACTGAGC 425

RESULT 11

AQ516886

LOCUS
 DEFINITION

RPCI-11-319E4.TV

genomic c

RPCI-11-319E4,
 genomic survey sequence.

ACCESSION

AQ516886

VERSION

1 GI:488576

GSS.

KEYWORDS

Homo sapiens (human)

ORGANISM

Homo sapiens

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Buteleostomi;

Metazoa; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 657)

REFERENCE

Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and

Venter,J.C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

PUBLISHED

1997

JOURNAL

COMMENT

Unpublished

1997

TITLE

COMMENT

Unpublished

1997

COMMENT

Unpublished</p

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM10538 row: 9 column: 14

High quality sequence stop: 757.

FEATURES

source

1..905

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4780059"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NCI CGAP Skn3"

/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.5kb. Library constructed by Life

Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match Similarity 79.1%; Score 18.2; DB 4; Length 905;

Best Local Similarity 87.0%; Pred. No. 7.3e+02; Mismatches 3; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy

1 GTGTCCTTAAGGAGACACCAAGC 23

Db

142 GTGTCCTTAAGGAGACACCAAGC 164

Search completed: March 9, 2005, 08:47:38

Job time : 3241 secs

FEATURES

source

1..1264

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3855873"

/tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.8 kb. Library constructed by Life

Technologies."

ORIGIN

Query Match Similarity 79.1%; Score 18.2; DB 2; Length 1264;

Best Local Similarity 87.0%; Pred. No. 7.6e+02; Mismatches 3; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

This Page Blank (uspto)

This Page Blank (uspto)

This Page Blank (uspto)